

DESIGN AND IMPLEMENTATION OF A DIGITAL HOLOGRAPHIC
MICROSCOPE WITH FAST AUTOFOCUSING

by
AYTEKİN HAZAR İLHAN

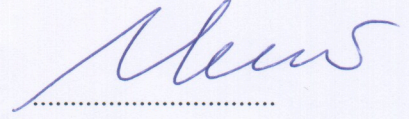
Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Master of Science

Sabancı University
Spring 2014

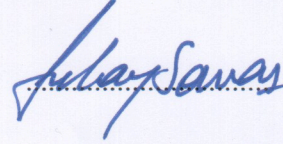
DESIGN AND IMPLEMENTATION OF A DIGITAL HOLOGRAPHIC
MICROSCOPE WITH FAST AUTOFOCUSING

APPROVED BY:

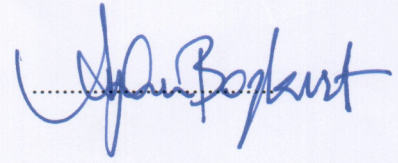
Assoc. Prof. Dr. Meriç Özcan
(Thesis Supervisor)



Assoc. Prof. Dr. Erkan Savaş



Assoc. Prof. Dr. Ayhan Bozkurt



DATE OF APPROVAL: 03/09/2014

© Aytakin Hazar İlhan 2014

All Rights Reserved

DESIGN AND IMPLEMENTATION OF A DIGITAL HOLOGRAPHIC MICROSCOPE WITH FAST AUTOFOCUSING

Aytekin Hazar İlhan

EE, M.Sc. Thesis, 2014

Thesis Supervisor: Assoc. Prof. Dr. Meriç Özcan

Keywords: interferometry, digital holography, digital holographic microscopy, computer generated holography, phase-shifting holography, sharpness, autofocusing, graphics processor

Abstract

Holography is a method for three-dimensional (3D) imaging of objects by applying interferometric analysis. A recorded hologram is required to be reconstructed in order to image an object. However one needs to know the appropriate reconstruction distance prior to the hologram reconstruction, otherwise the reconstruction is out-of-focus. If the focus distance of the object is not known priori, then it must be estimated using an autofocusing technique. Traditional autofocusing techniques used in image processing literature can also be applied to digital holography. In this thesis, eleven common sharpness functions developed for standard photography and microscopy are applied to digital holograms, and the estimation of the focus distances of holograms is investigated. The magnitude of a recorded hologram is quantitatively evaluated for its sharpness while it is reconstructed on an interval, and the reconstruction distance which yields the best quantitative result is chosen as the true focus distance of the hologram. However autofocusing of high-resolution digital holograms is very demanding in means of computational power. In this thesis, a scaling technique is proposed for increasing the speed of autofocusing in digital holographic applications, where the speed of a reconstruction is improved on the order of square of the scale-ratio. Experimental results show that this technique offers a

noticeable improvement in the speed of autofocusing while preserving accuracy greatly. However estimation of the true focus point with very high amounts of scaling becomes unreliable because the scaling method detracts the sharpness curves produced by the sharpness functions. In order to measure the reliability of autofocusing with the scaling technique, fifty computer generated holograms of gray-scale human portrait, landscape and micro-structure images are created. Afterwards, autofocusing is applied to the scaled-down versions of these holograms as the scale-ratio is increased, and the autofocusing performance is statistically measured as a function of the scale-ratio. The simulation results are in agreement with the experimental results, and they show that it is possible to apply the scaling technique without losing significant reliability in autofocusing.

YÜKSEK HIZDA OTOMATİK ODAKLAMA YAPABİLEN SAYISAL HOLOGRAFİK MİKROSKOP TASARIMI VE GERÇEKLENMESİ

Aytekin Hazar İlhan

EE, Yüksek Lisans Tezi, 2014

Tez Danışmanı: Doç. Dr. Meriç Özcan

Anahtar Kelimeler: interferometre, dijital holografı, dijital holografik mikroskopi, bilgisayarda üretilmiş holografı, faz-değişimi ile holografı, netlik, otomatik odaklama, grafik işlemcisi

Özet

Holografı, nesneleri girişim desenleri ile analiz ederek üç boyutlu görüntü elde edebilmeyi sağlayan bir metoddur. Ancak bir hologramı yapılandırabilmek için nesnenin kamera düzleminden uzaklığının bilinmesi gereklidir, aksi takdirde elde edilen yapılandırmalar odaklanmamış olacaktır. Eğer bu 'yeniden yapılandırma uzaklığı' önceden bilinmiyorsa, bu mesafenin otomatik odaklama tekniği kullanılarak hesaplanması gerekir. Görüntü işleme literatüründe bulunan geleneksel otomatik odaklama teknikleri sayısal holografıye uygulanabilmektedir. Bu tezde, on bir geleneksel netlik kriteri dijital hologramlara uygulanarak, doğru yapılandırma mesafesinin bulunması incelenmiştir. Kaydedilen hologramlar çeşitli uzaklıklarda yapılandırılarak elde edilen görüntülerin genlikleri netlik kriterleri ile karşılaştırılmıştır. Sayısal olarak en keskin hatlara sahip genlik resmi, odaklanmış olan yeniden yapılandırmayı ifade etmektedir. Bu şekilde elde edilen yeniden yapılandırma uzaklığına 'doğru-odak-uzaklığı' ismi verilmiştir. Ancak, yüksek çözünürlükteki hologramların otomatik odaklanması oldukça uzun sürmektedir. Bu amaçla bir ölçekleme metodu geliştirilmiş ve sunulmuştur. Bu ölçekleme metodu ile doğru-odak-uzaklığı halen yüksek hassasiyet ile hesaplanabilirken, hesaplama süresi de ölçeğin karesi oranında kısalmaktadır. Ancak, çok yüksek ölçek değerleri kullanıldığında otomatik

odaklama işlemi güvenilir olmamaktadır, çünkü ölçekleme metodu netlik kriterleri ile hesaplanan netlik eğrilerinde bozulmaya yol açmaktadır. Bu sebeple, ölçekleme metodu ile otomatik odaklama işleminin güvenilirliğini ölçmek için insan portresi, manzara ve mikro-yapılar içeren elli adet resimden sentetik hologramlar türetilmiştir. Bu hologramlar artan ölçek değerleri ile ölçeklenerek otomatik odaklamaya tabi tutulmuş, ve otomatik odaklamanın güvenilirliği istatistiksel olarak incelenmiştir. Simülasyon sonuçları deneysel sonuçlar ile uyuşmakta, ve ölçekleme tekniği kullanılarak güvenilir otomatik odaklama yapılabileceği gösterilmektedir.

To my beloved family

ACKNOWLEDGMENTS

I would like to thank my thesis advisor Prof. Dr. Meriç Özcan for his continuous support and guidance through both my graduate and undergraduate studies. It has been a privilege studying under his guidance. I would like to thank to my thesis committee members; Prof. Dr. Erkey Savaş and Prof. Dr. Ayhan Bozkurt for their interest and constructive criticisms. I thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) for funding this project under the research fund No. 110T613. I would also like to thank TÜBİTAK again for supporting me with BİDEB-2210 scholarship throughout my graduate education.

I would like to thank to all my friends who encouraged and supported me in my years in Sabancı University. Special thanks to Mert Doğar, for we have overcome numerous sophisticated projects together with him, and his level of dedication always inspired me.

And last but not least, I would like to thank to my family for their endless love and support in all my life. A special mention goes to my grandparent Şükrü Aytekin İlhan, greatly missed, who always believed in me and mediatly enabled my part in this research.

TABLE OF CONTENTS

Abstract	iv
Abstract (Turkish)	vi
Acknowledgments	ix
Table of Contents	x
List of Figures	xii
List of Tables	xviii
1 INTRODUCTION	1
2 DIGITAL HOLOGRAPHY	5
2.1 Digital Holographic Recording	5
2.2 Digital Hologram Filtering Methods	9
2.2.1 Spatial Filtering	9
2.2.2 Phase-Shifting Holography	11
2.3 Digital Hologram Reconstruction Methods	13
2.3.1 Reconstruction using Fresnel approximation	14
2.3.2 Reconstruction using convolution	17
2.3.3 Reconstruction using angular spectrum	18
2.4 Computer Generated Holograms	20
3 DIGITAL HOLOGRAPHIC MICROSCOPE	22
3.1 Optical Setup	22
3.2 Integration of the Graphics Card to the DHM	30
3.3 Control Interface of the DHM	32
3.4 Hologram Reconstruction Experiments with the DHM	34
4 AUTOFOCUSING	36
4.1 Focusing of Digital Holograms	36
4.2 Autofocusing in Digital Holography	37
4.2.1 Sharpness Metrics	39
4.3 Experimental Results on Autofocusing	44

5	METHODS FOR FAST AUTOFOCUSING	47
5.1	Intelligent Search Algorithms	48
5.2	Utilizing the Graphics Processing Unit (GPU)	49
5.3	Autofocusing using The Scaling Technique	50
5.3.1	Autofocusing Experiments using the Scaling Technique	53
5.3.2	Statistical Investigation of the Scaling Technique on Computer Generated Holograms	57
6	CONCLUSION	62

LIST OF FIGURES

2.1	An off-axis hologram of Lena's image is presented. When the hologram is reconstructed without filtering, four sub-images are revealed. In (b), the image on the upper right corner is the virtual object image, the image on the lower left corner is the diffracted twin image, and the bright blurry image in the middle is the composition of the zero-order images.	8
2.2	An on-axis hologram of Lena's image, where the four sub-images are superimposed on each other. In (b), the image of Lena is not clearly visible due to the distortion effects caused by the zero-order and the twin images.	9
2.3	Spatial filtering is demonstrated on the hologram of Lena's image. (a) The off-axis hologram of Lena. (b) 2D frequency spectrum of the hologram. (c) The filtered frequency spectrum. (d) The hologram in spatial domain after filtering. (e) Reconstruction of the filtered hologram.	10
2.4	Filtering by the phase-shifting technique is demonstrated on an off-axis hologram of Lena's image. (a) The off-axis hologram of Lena. (b) The reconstruction after filtering zero-order terms by using an additional hologram. (c) The reconstruction after filtering the zero-order and twin images by using two additional holograms.	13

3.1	A Mach-Zehnder interferometer based digital holographic microscope setup that can perform phase-shifting holography. The laser beam is divided into two arms by a beam-splitter (BS1). The object wave is transmitted through the specimen (S) while the phase of the reference wave is modulated by a phase-modulator (PM). The phase-modulator is connected to a high-voltage amplifier which is connected to National Instruments USB-6251 that is controlled by a personal computer. The beams are magnified by identical microscope objectives and they are superimposed on each other using another beam-splitter (BS2). The formed interference pattern is recorded by a CCD camera that is connected to the personal computer.	23
3.2	Digital holographic microscope. The laser beam is animated for better illustration.	24
3.3	National Instruments USB 6251 (NI) is used for driving the optoelectronic phase modulator. The device is connected to the personal computer via USB cable. After the desired voltage information is fed from the computer digitally, NI applies that much voltage from its analog port A01 to a high-voltage amplifier which is connected to the phase modulator. . . .	25
3.4	Two identical microscope objectives (MO) are placed on the path of the object and the reference beam for equalizing the wavefront curvatures of the waves. The microscope objectives are symmetrical with respect to the beam-splitter in x, y and z axes.	25

3.5	The optical components of the transmission based digital holographic microscope (DHM) setup are shown in correlation with the schematic. The laser beam is divided into two arms by a beam-splitter (BS1). The reference wave is modulated by a phase modulator (PM) and the object wave is transmitted through the specimen. Before joining the two arms of the interferometer by a second beam-splitter (BS2), object image is magnified by the microscope objective (MO1). In order to compensate the wavefront curvatures, a similar objective (MO2) is placed on the reference beam path. Finally, interference of the two beams is recorded by a CCD camera that is connected to a personal computer.	26
3.6	Snapshots of the digital holographic microscope.	28
3.7	The digital holographic microscope is enclosed within a black box for reducing the effects of vibrations in the air.	29
3.8	The calculation of the transfer function of the angular spectrum method implemented in CUDA language for execution in GPU	30
3.9	The reconstruction procedure of a recorded hologram using the graphics processing unit.	31
3.10	A snapshot of the controlling interface of the software developed for operation of the DHM.	33
3.11	A snapshot of the display window of the software developed for the DHM.	33
3.12	A hologram of onion cells is recorded 3 cm away from the recording plane. The magnitude (top row) and the phase (bottom row) of the holographic reconstructions by the a) Angular spectrum, b) Fresnel approximation and c) Convolution methods are presented.	34
3.13	A hologram of USAF resolution chart is recorded 4.5 cm away from the recording plane. The magnitude (top row) and the phase (bottom row) of the holographic reconstructions by the a) Angular spectrum, b) Fresnel approximation and c) Convolution methods are presented.	35

3.14	A hologram of USAF resolution chart is recorded 1 cm away from the recording plane. The magnitude (top row) and the phase (bottom row) of the holographic reconstructions by the a) Angular spectrum, b) Fresnel approximation and c) Convolution methods are presented.	35
4.1	Autofocusing is shown on a computer generated hologram of the USAF chart. The hologram is reconstructed between the interval 0 to 6 cm with 0.1 mm step-size. Then each reconstruction is evaluated by the normalized-variance metric and the sharpness curve is formed. Finally, the hologram is reconstructed at the propagation distance that corresponds to the peak of the sharpness curve.	38
4.2	(a) and (c) are the recorded holograms of USAF resolution chart and the photoresist test subject. (b) and (d) are the reconstructions of the holograms at the exact focus distances 3 cm and 1.1 cm respectively.	45
4.3	Sharpness curves obtained by traditional focus metrics using the recorded holograms of USAF resolution chart and the photoresist test subject. The true focus distances of the holograms are 3 cm and 1.1 cm respectively. The sharpness curves are calculated on a 5 cm interval with a 0.5 mm step size. The sharpness curves of the USAF chart are represented by continuous lines, and the sharpness curves of the phase object are represented by dotted lines. Tenenbaum gradient, the integral power, the normalized-variance and the deviation-based correlation metrics performed better. . .	46
5.1	Scaling operation is demonstrated. A hologram is divided into k by k squares, and the average intensity of each square is set as the corresponding pixel value of the new scaled hologram.	51

- 5.2 An exemplary autofocus using the scaling technique. A recorded hologram of the USAF chart (a) is scaled with ratio $k = 16$ and the scaled-hologram (b) is obtained. Then the sharpness curve of this new scaled hologram (c) is computed using the normalized-variance and the gradient-squared metrics, represented by continuous and dotted lines respectively. The true-focus-distance of the scaled hologram is identified as 3 cm by both of the sharpness metrics as a global maxima. Finally, the original sized hologram is reconstructed at true-focus-distance and a focused image of the object is achieved (d). 52
- 5.3 The effect of scaling on sharpness curves is shown on a recorded hologram of USAF resolution chart. Hologram is originally composed of 1024×1024 pixels. Two, four, eight, 16 and 32 times scaling is applied and the sharpness curves are calculated by the four chosen metrics. As k , increases curves mostly become wider and smoother due to the low-pass nature of the scaling. 53
- 5.4 (a) Hologram of human cheek epithelial cells, and (b) its reconstruction at the true focus distance. The magnitude of the reconstructed holograms are used for finding the true focus distance (2.4 cm in this case), where the epithelial cells are normally invisible in the image when it is focused (b - top). After finding the focus distance, unstained epithelial cells are visible in detail in the phase image of the reconstruction (b - bottom). To find the focus distance, the hologram is scaled two, four, eight, 16 and 32 times, and the sharpness curve of each scaled version is computed using the normalized-variance, the integral power, Tenenbaum gradient and the deviation-based correlation metrics. The change in the resolution and the peak-point deviation of the sharpness curves obtained by these four methods are shown in (c) and (d) respectively for increasing scale-ratios. As the resolution of the curves decrease, it is possible to scale up to 8 times safely without diverging from the actual focus distance 56

5.5	The list of fifty different 1024×1024 pixels gray-scale images that are used for creating computer generated holograms.	59
5.6	Fifty different computer generated holograms of 1024×1024 pixels are subjected to sharpness estimation using the normalized-variance, the integral power, Tenenbaum gradient and the deviation-based correlation metrics. For each scale-ratio, computer generated holograms are scaled, and then the sharpness curves are calculated by each method in 100 steps between 22 cm and 28 cm. For each metric, the curves obtained from fifty images are overlaid on each other as a function of the scale-ratio.	60
5.7	Fifty different computer generated holograms of 1024×1024 pixels are subjected to sharpness estimation using the normalized-variance, the integral power, Tenenbaum gradient and the deviation-based correlation metrics. For each scale-ratio, computer generated holograms are scaled, and then the sharpness curves are calculated by each method in 100 steps. (a) Average sharpness curve quality (FWHM^{-1} of the global extrema) for each metric as a function of scale-ratio. (b) Average deviation of the global extrema from the true focus distance for each metric as a function of scale-ratio. Note that the normalized-variance and the deviation-based correlation metrics produced some flipped and unreliable sharpness curves when $k \geq 16$. For example, for $k = 32$ a single curve of the normalized-variance metric is flipped. Those flipped curves are discarded while calculating the average performance of the metrics.	61

LIST OF TABLES

3.1	Specifications of the digital holographic microscope.	27
4.1	The description an the mathematical representation of the derivative based focus metrics used in this work are listed, where $O_{x,y}$ indicates the magnitude of the object image at pixel location (x, y)	40
4.2	The description an the mathematical representation of the histogram based focus metrics used in this work are listed, where $O_{x,y}$ indicates the magnitude of the object image at pixel location (x, y)	41
4.3	The description an the mathematical representation of the power based focus metrics used in this work are listed, where $O_{x,y}$ indicates the magnitude of the object image at pixel location (x, y)	42
4.4	The description an the mathematical representation of the statistics based focus metrics used in this work are listed, where $O_{x,y}$ indicates the magnitude of the object image at pixel location (x, y)	43
5.1	The time consumed by a hologram reconstruction using angular spectrum method and the time consumed by a sharpness estimation using normalized-variance and gradient-squared metrics are shown. Hologram reconstruction is completed much more slower than the sharpness estimation, and thus it is the bottleneck in an autofocusing algorithm. Note that the timings are obtained with an Intel i7-2600 CPU working on a single core.	48

5.2	Reconstruction times in milliseconds for the hologram reconstruction methods. 1024×1024 holograms are reconstructed by an Intel XEON W3670 CPU and an Nvidia GTX 660 Ti graphics card.	49
5.3	The figure shows the execution time of a single hologram reconstruction using the convolution method and a single execution of the normalized-variance metric. The timings are obtained on an Intel i7-2600 CPU. . . .	50

Chapter 1

INTRODUCTION

Holography is invented by Dennis Gabor in 1948 while he was working on electron microscopy [1]. He recorded the interference pattern of a wave scattered from an object and a background wave on a photographic film, and then he showed that it was possible to reconstruct the object wave-field by illuminating the photographic film with the reference wave that is used in the recording. With this technique, he was able to recover both the intensity and phase information of the original object wave-field. He coined his work as 'wavefront reconstruction' at first, but later he addressed the technique as 'holography'. The word holography is inspired from two Greek words 'holos' and 'graphen' which mean 'whole' and 'to write' respectively.

The first holograms recorded by Gabor had important problems such as the weak coherence between the waves, and the presence of a twin-image of the object. Ten years after his work in 1959, Gould has invented a highly coherent light-source, the laser, which induced a rapid development rate in interferometric applications [2]. Just three years later, the twin-image problem is addressed by Leith and Upatnieks who invented off-axis holography where there is a small angle between the object and the reference waves [3]. With their setup, the twin images were spatially separated. In the late 1960s, holographic interferometry is developed by Powell and Stetson, and it started to have use in applications such as vibration analysis and surface contouring [4, 5, 6].

In 1969, Brown and Lohmann introduced the concept of generating holograms in a computer environment, and then reconstructing them optically [7]. This technique is

called computer generated holography (CGH), and the process of synthesizing holograms is explored by other scientists as well [8, 9]. A few years later, in the early 1970s, process of hologram reconstruction is transferred into the computer environment [10, 11, 12]. This gave birth to the idea of digital holography in which the optical illumination of a hologram with the reference wave and the wave propagation procedures are performed numerically on a computer [13]. In 1993, Schnars and Jüptner established the idea of digital holography is established where they recorded holograms digitally on charge coupled devices (CCDs) [14, 15]. With their method, the photographic recording was removed as an intermediate step, and it became possible to directly record and reconstruct holograms numerically. Since the mid 1990s, the topic of digital holography became very popular, and it has started to have use in many applications such as deformation analysis and shape measurement [16, 17], microscopy [18, 19, 20, 21], particle measurement and tracking [22, 23, 24], refractive index measurement [25, 26], holographic data storage [27, 28, 29] and watermarking [30].

Digital holographic imaging has three significant advantages over standard imaging techniques. First of all, since the phase of the object wave is coupled to the recorded intensity image due to interference, both the magnitude and phase of the object wave can be recovered numerically [31], which allows imaging of phase objects without any staining. Secondly, an exact replica of the object wave-field is recovered in holographic imaging, and therefore it is considered as the only true 3D imaging technique compared to other imaging technologies such as stereoscopic imaging or integral imaging [32, 33, 34]. Thirdly in holographic imaging, no mechanical focusing is required at the hologram recording step, because the recorded hologram can be numerically reconstructed at any desired depth. However to obtain a focused 3D image of a specimen, the hologram must be reconstructed at the actual depth of the object.

When the actual depth of an object is not known a priori, the hologram is reconstructed at several different depths and the best focused reconstruction is hand-picked by a human. An autonomous solution to this problem is using an autofocus algorithm for selecting the best focused object image. There are numerous metrics proposed in literature for comparing the in-focusness of images acquired by photography and light-

microscopy [35, 36, 37], where the stage or the lens position is adjusted successively as the images are evaluated by a sharpness metric. The very same approach is valid in digital holography and digital holographic microscopy as well. Basically, the hologram is reconstructed at candidate distances and the sharpness of the magnitude of the reconstructed images are compared using a sharpness metric. Although, there are sharpness metrics explicitly developed for digital holography in literature [38, 39, 40], the traditional sharpness metrics are able to estimate the true-focus-distances of holograms accurately as well [41].

Autofocusing of digital holograms may be very demanding in means of computational power when the hologram size is large. For example, a one mega-pixels size hologram is reconstructed in hundreds of milliseconds with a high-end conventional Central Processing Unit (CPU) [42]. In order to achieve real-time operation speed, hardware accelerators such as Graphics Processing Units (GPUs) may be utilized as the calculation engine [43, 44, 45]. However this comes with a capital cost for the hardware, and an engineering cost for integrating it to the system. Another way to achieve faster operation is using scaled holograms for finding out the true-focus-distance of the original hologram [46]. When a hologram is scaled-down k times in size, its reconstruction is completed approximately k^2 times faster than the original. It is worth to note that the speed improvement using scaling technique is independent of the processing power of the system, and the speed can be further improved by many-folds if GPUs were to be utilized in addition.

In this thesis, autofocusing is performed on digital holograms recorded by a DHM using traditional sharpness metrics, and the performance of the scaling method is presented. In chapter 2, principles of digital holographic recording, hologram filtering and hologram reconstruction methods are given. In chapter 3, the specifications of our digital holographic microscope and the developed software for controlling the DHM are presented. Afterwards, exemplary in-focus reconstructions of captured holograms are shown. In chapter 4, autofocusing in digital holography is investigated and eleven of the most common sharpness metrics are presented to be used in autofocusing. Then, the sharpness functions are tested experimentally, and four of them are chosen to be the most suited for autofocusing in digital holography. In chapter 5, methods for increasing the speed perfor-

mance of autofocusing are explored, and the scaling technique is investigated. After validating the technique experimentally, the amount of degradation due to scaling in means of accuracy is statistically investigated on fifty computer generated holograms (CGH), and the results are presented. Finally in the last chapter, the work done in the thesis is summarized and the conclusion is provided.

Chapter 2

DIGITAL HOLOGRAPHY

In digital holography, the interference pattern of a wave scattered from an object and a reference wave is recorded with a CCD camera, and the object image is reconstructed in 3D [47, 31]. Since the phase of the object wave is coupled to the recorded intensity image due to interference, both the magnitude and phase of the object wave can be recovered numerically.

In this chapter, principles of digital holographic recording and filtering of digital holograms are explained. Three different methods for reconstructing digital holograms are described and their performances are discussed. Finally, a method for creating computer generated holograms is described.

2.1 Digital Holographic Recording

A holographic recording makes use of the interference of two coherent light fields, where one field is regarded as the object wave O and the other one is regarded as the reference wave R . One of the easiest ways to achieve such a configuration is by dividing a laser beam into two arms by use of a beam-splitter. Then, one arm is used to illuminate the object and the wave diffracted from the object is recorded by the recording medium. On the other hand, the other arm is object-free and the beam directly hits the recording medium.

In a digital holography setup, the recording medium is typically a CCD/CMOS camera. When an image is shot with a camera, only the intensity of the light field is captured.

For example, if an object is illuminated by a beam and the beam that is scattered from the object is O , then the recorded image I is expressed as

$$I = |O_o|^2. \quad (2.1)$$

The phase information is lost in this process, and there is no known way to recover the 3-D properties of O without using some extra information or utilizing extra images.

However in digital holography, the intensity of summation of the object and reference waves are recorded by the camera. Since these two light waves are coherent and they interfere along the optical axis, the recorded image contains fringes that gives information about the 3-D structure of the object. The wave scattered from the object (O) and the reference wave (R) on the recording medium is then expressed as [15]:

$$I = |O + R|^2. \quad (2.2)$$

In this type of recording, I is no longer a simple intensity image, and instead it is called as the "the hologram of the object". The hologram I can be expanded mathematically, and four sub-images are revealed inside of it:

$$I = |O|^2 + |R|^2 + OR^* + O^*R. \quad (2.3)$$

Each of the four images has their own characteristics. The first two images are called the zero order images of the object and the reference waves. These two images are undesirable for holographic imaging because they distort the others. The first term $|O|^2$ is a standard image of the object. Second term $|R|^2$ is a spatially invariant DC signal which increases the overall intensity of the hologram. There are methods to get rid of those two disturbances, which will be explained shortly. The third and the fourth mathematical terms represent the object image multiplied with the reference wave and the complex conjugate of this object image. These two images carry the required information to reconstruct the object wave completely in 3D.

To reconstruct a hologram, it must be illuminated by the same reference wave that was used in the recording process. Commonly, a plane reference wave is used in the recording process to ease up the reconstruction. The illuminated hologram is mathematically

expressed as:

$$I.R = |O|^2 R + |R|^2 R + OR^* R + O^* R R. \quad (2.4)$$

When the reference wave is a plane wave, the multiplications with the reference wave are simplified like a multiplication with a constant term and a spatially invariant phase shift on the multiplied image. The zero order terms diffract in the illumination direction and since the object image O is spatially variant, it causes distortions on the rest of the image. The third and the fourth images are the virtual and the conjugate object waves which diffract to opposite sides along the z -axis. Although the planar reference wave assumption simplifies the reconstruction, this is not required, and as long as the characteristics of the reference wave is known, it is possible to reconstruct the hologram.

The four sub-images inside a hologram are best separated visually when the recording is performed off-axis. Off-axis recording is characterized by a slight angular difference between the object beam and the reference beam. Figure 2.1-a shows an off axis hologram of Lena's image where the object and reference beams are slightly off in x and y axes. Note that the resolution of the hologram is directly dependent on this angular difference and the more the difference the lesser the resolution. Another result of this angular difference is that the four sub-images are located on physically different portions of the ccd when the hologram is reconstructed. When the hologram is illuminated by the reference wave, the image in Figure 2.1-b is obtained. This image clearly shows the four sub-images located in a hologram. Note that these images are also physically located at different depths on z -axis such that they come in to focus at different depths.

When an off-axis hologram is reconstructed, an image similar to Figure 2.1-b is acquired. The bright square in the middle of the image is the composition of the zero order images of the hologram. These zero order images do not diffract to sides and they remain in the middle. The image of Lena on the upper right corner is the virtual image that was reconstructed. The diffracted image on the lower left corner is the twin image of the object and it is not focused. To focus the twin image instead, it is required to illuminate the hologram from the opposite direction (on z axis). The twin images diffract further into spatially opposite corner sides as the reconstructing beam travels further.

Although it is easy to get rid of the zero-order images and the conjugate image in

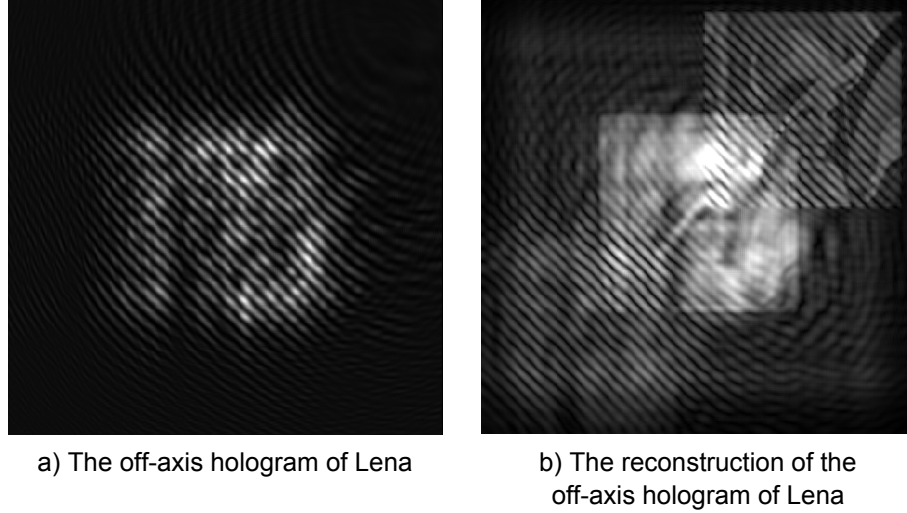
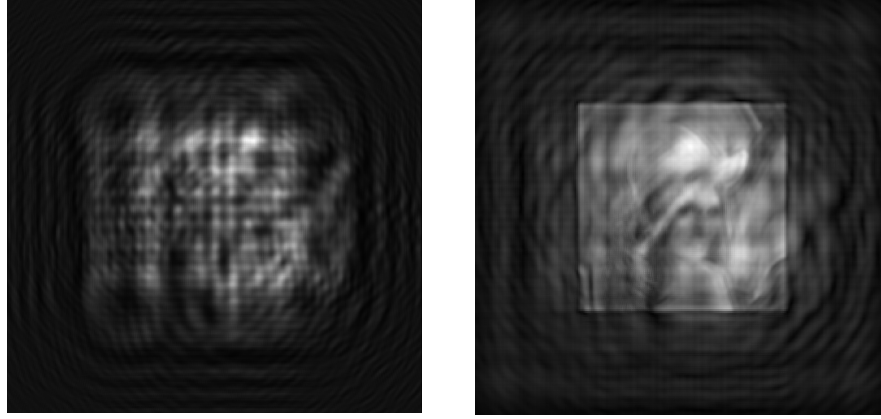


Figure 2.1: An off-axis hologram of Lena's image is presented. When the hologram is reconstructed without filtering, four sub-images are revealed. In (b), the image on the upper right corner is the virtual object image, the image on the lower left corner is the diffracted twin image, and the bright blurry image in the middle is the composition of the zero-order images.

off-axis recording, this type of recording is not very desirable because there is a major drawback. In off-axis recording, most of the ccd array is not used for imaging and ccd pixels are wasted, for example, only the upper right corner of the imaging field is used for viewing Lena's image in Figure 2.1. As one would recognize, the lesser the angle between the object and reference beams, the more the number of ccd pixels that are used for imaging and not being wasted. This leads us to the in-line holography where there is no angle ($\theta = 0$) between the two arms.

In in-line holography, the four sub-images cannot be separated visually, because all the sub-images are centered in x and y axes, and they are superimposed on each other. However, one would like to acquire only the third term in equation 2.4 -that is O^*RR -. The first term $|O|^2R$ causes self interference, the second term $|R|^2R$ adds a DC intensity offset and the fourth term O^*RR causes distortions. In figure 2.2, an in-line hologram and its reconstruction is presented where the actual image of the object is distorted and not visible due to the other images in the hologram. To image the object as desired, one needs to filter out the zero order images and the unfocused twin image. When this is achieved, the ccd pixels are fully utilized for imaging and the resolution is the highest.



a) An on-axis hologram of Lena's image

b) The reconstruction of the hologram of Lena's image

Figure 2.2: An on-axis hologram of Lena's image, where the four sub-images are superimposed on each other. In (b), the image of Lena is not clearly visible due to the distortion effects caused by the zero-order and the twin images.

2.2 Digital Hologram Filtering Methods

In holographic context, filtering means getting rid of these undesired images in the reconstruction, and different filtering methods exist for in-line (on-axis) and off-axis methods. While a simple spatial filtering in the frequency domain is sufficient in off-axis holography, in-line holography requires a more advanced operation with the capture of multiple holograms of a scene. In the following, the two filtering techniques are explained.

2.2.1 Spatial Filtering

Spatial filtering is a simple way to suppress the undesirable images in an off-axis hologram. This technique depends on the separation angle between the object and the reference beam. For example mathematically, when the reference beam is off with respect to the object beam in x axis, the recorded pattern is expressed as [48]:

$$I = |O|^2 + |R|^2 + ORe^{-ikx\sin\theta} + O^*Re^{ikx\sin\theta}, \quad (2.5)$$

where k is the wave number and θ is the angle between the reference beam and the object beam in x axis. The phase expressions on the third and the fourth images represent the result of the angular difference between the beams. This difference provides a spatial shift to the twin images in the frequency spectrum of the hologram I when 2D Fourier

transform is applied. Figure 2.3-a and 2.3-b shows an off-axis hologram of Lena's image and its frequency spectrum. In this example, the reference and the object beams are off in both x and y axes, therefore the frequency components of the twin images reside at the upper-right and lower-left corners. On the other hand, the frequency components of the zero-order images remain at the center as bright squares. For an angle θ on the reference beam, the frequency components of the twin images are shifted to $(-k\sin\theta/2)$ and $(k\sin\theta/2)$ locations in the spectrum.

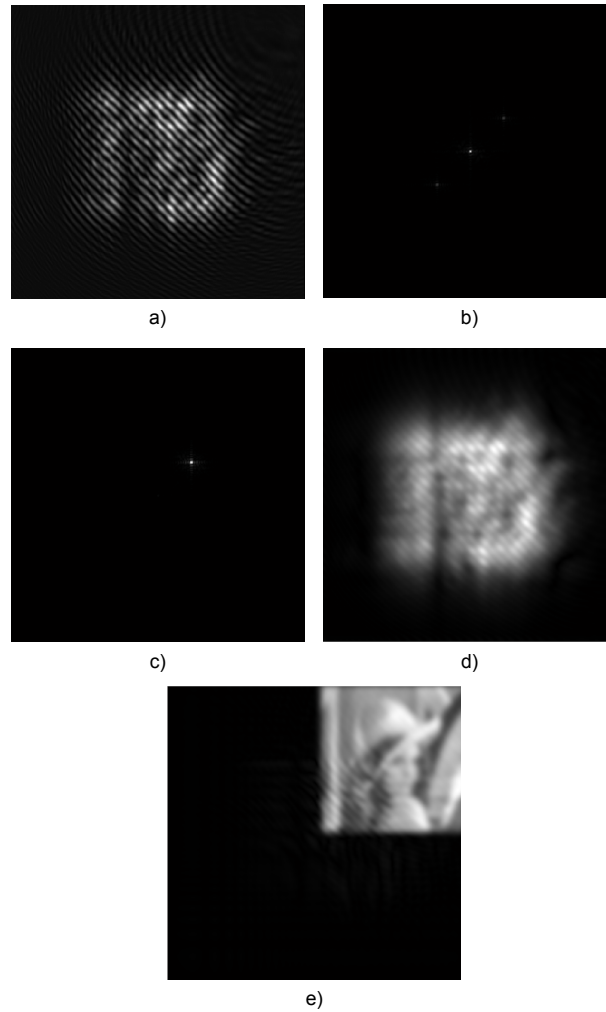


Figure 2.3: Spatial filtering is demonstrated on the hologram of Lena's image. (a) The off-axis hologram of Lena. (b) 2D frequency spectrum of the hologram. (c) The filtered frequency spectrum. (d) The hologram in spatial domain after filtering. (e) Reconstruction of the filtered hologram.

Using a two dimensional Gaussian band-pass filter, it is possible to isolate any of the twin images [15]. In figure 2.3-a, the hologram of Lena's image is presented and in

figure 2.3-b, the 2D frequency spectrum of the hologram is computed. Then, frequency components of the virtual image is isolated by a Gaussian filter in figure 2.3-c. Afterwards, the inverse Fourier transform is applied to obtain the filtered hologram as shown in figure 2.3-d. Finally, figure 2.3-e shows the reconstructions of the filtered hologram where spatial filtering is applied. As it is seen, the reconstructions do not contain the zero-order images and it is possible to isolate the virtual image alone, however this at the expense of resolution of the hologram.

The problem with spatial filtering is that it requires a angle between the object and reference beams, and therefore this method is inherently not compatible with on-axis holography where there is no angle between the arms by definition. Moreover greater the angle θ , lesser the resolution of the hologram. Therefore, other techniques are developed for filtering holograms other than by spatial filtering, such as the phase-shifting holography.

2.2.2 Phase-Shifting Holography

Phase-shifting holography is a technique for filtering a hologram by mathematically combining multiple captures of a scene [49]. For each additional capture of a scene, the phase of the reference beam is set to a different level precisely, so that the interference pattern shifts slightly. This yields out valuable information and makes possible to remove zero-order images completely, or isolate the virtual object image.

Since phase-shifting method does not depend on the angle between the arms, it is a viable filtering solution for both on-axis and off-axis holography. However the method depends on the time-invariance of the scene that is being captured, and while this does not pose a problem for still objects, it is an important consideration for objects with velocity or for alive specimens, and thus the required equipment should be chosen appropriately.

To perform phase-shifting holography, it is required to introduce a physical phase-shifter into the holographic setup. The simplest option is to add an opto-electronic phase modulator into the optic configuration. This phase-modulator should be set on the path of the reference wave only. Modulating the phase of light wave with Mhz rates, it is possible to produce phase-shifted holograms very fast and without dealing with time-

variance. Another low-cost option for using a phase-shifter is modulating the position of a mirror on the path of the reference beam with nanometer precision. However this adds an engineering cost and requires precise calibration, and it must be guaranteed that the mirror is stabilized faster than the time-variance in the scene. Finally, the capture rate of the CCD camera should also match with the equipment, and its frame-rate must also be greater than the amount time lapsed before the scene varies.

Filtering zero-order images with two captures

To filter out the zero-order images, only 1 additional capture of a scene is sufficient. Initially, the recorded hologram is:

$$I_0 = |O + R|^2 = |O|^2 + |R|^2 + OR^* + O^*R. \quad (2.6)$$

The phase-shifted version of the hologram is obtained by introducing a π shift to the phase of the original reference wave. Then the second hologram is:

$$I_\pi = |O + Re^{j\pi}|^2 = |O|^2 + |R|^2 - OR^* - O^*R. \quad (2.7)$$

To obtain the zero-order free hologram, it is sufficient to subtract the holograms from each other. That is:

$$I_0 - I_\pi = 2(OR^* + O^*R). \quad (2.8)$$

Isolating the real object image with three captures

By algebraically introducing a third phase-shifted hologram of the same scene, it is possible to remove all the undesirable images from a hologram. Suppose a third capture with $\frac{\pi}{2}$ shift on the phase of the reference wave is recorded, that is:

$$I_{\frac{\pi}{2}} = |O + Re^{j\frac{\pi}{2}}|^2 = |O|^2 + |R|^2 - jOR^* + jO^*R. \quad (2.9)$$

Then to obtain the real object image alone, one needs to combine all the three holograms with the following formula:

$$OR^* = \left[\frac{1}{4}(I_0 - I_{\pi/2}) + j\frac{1}{4}(2I_{\pi/2} - I_\pi - I_0) \right]. \quad (2.10)$$

Multiplying each side of the equation with R , one directly obtains the object wave O :

$$O = \left[\frac{1}{4}(I_0 - I_{\pi/2}) + j\frac{1}{4}(2I_{\pi/2} - I_{\pi} - I_0) \right] R. \quad (2.11)$$

The amplitude of the reference wave is taken as unity in the calculations because R is generally a uniform plane wave. This plane wave preference is important because it simplifies both the filtering and the reconstruction procedures in hologram processing as explained in the following section.

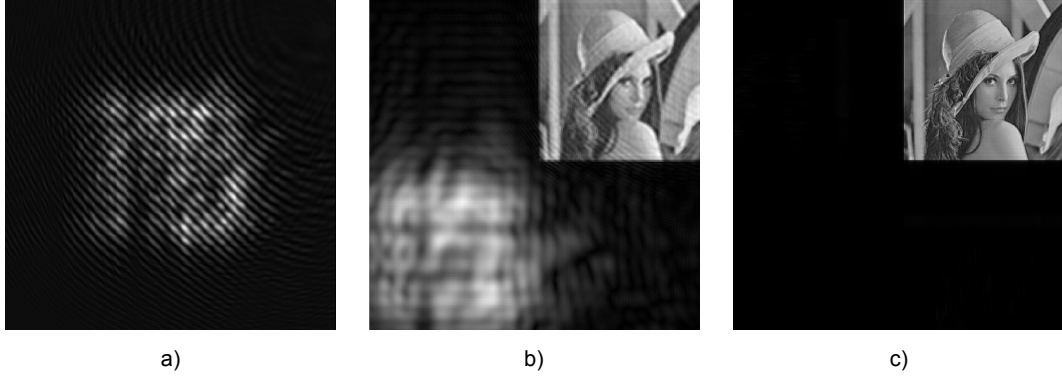


Figure 2.4: Filtering by the phase-shifting technique is demonstrated on an off-axis hologram of Lena's image. (a) The off-axis hologram of Lena. (b) The reconstruction after filtering zero-order terms by using an additional hologram. (c) The reconstruction after filtering the zero-order and twin images by using two additional holograms.

2.3 Digital Hologram Reconstruction Methods

In digital holography, a recorded hologram is reconstructed by numerically illuminating the hologram with the same reference wave that was used in the recording [50, 31]. This is achieved by modeling the reference wave numerically, then multiplying it with the hologram, and then finally numerically propagating the illuminated hologram from camera plane to an observation plane. Before proceeding further, it is important to note that the uniform planar reference wave assumption greatly simplifies this operation here, because then the amplitude of the reference wave is taken as 1. If the recorded hologram is in-line, the space-invariant phase-shift is further ignored, which means that modeling of the reference wave is skipped while illuminating the hologram, and the hologram is equal to itself

after the illumination. Numerical illumination of the hologram with the reference wave is mathematically performed by:

$$\begin{aligned} I_0(x, y) &= I_{recorded}(x, y)R(x, y), \\ R(x, y) &= 1.e^{ik \cdot z}, \\ I_0(x, y) &= I_{recorded}(x, y), \end{aligned} \tag{2.12}$$

where $I_{recorded}$ is the hologram before illumination, I_0 is the hologram after illumination, and R is a planar reference wave traveling in z direction with wave number k . Throughout this thesis, it is assumed that this planar wave assumption for R holds. If the planar wave assumption does not hold for a holographic recording configuration, then it is required to record the reference wave separately for numerical illumination.

The heart of the hologram reconstruction process is the numerical propagation of the illuminated hologram to an observation plane where the wave is focused -which is called the object plane-. To calculate the diffracted wave from camera plane to object plane, Rayleigh-Sommerfeld diffraction integral can be used [31]:

$$I_d(u, v) = \frac{1}{i\lambda} \iint I_0(x, y) \frac{\exp(-i\frac{2\pi}{\lambda}\rho)}{\rho} \cos \theta \, dx \, dy, \tag{2.13}$$

$$\rho = \sqrt{d^2 + (x - u)^2 + (y - v)^2}, \tag{2.14}$$

where I_0 is the distribution of the wave on the camera plane, ρ is the cartesian distance between the points of the camera and object planes, λ is the wavelength of the light, and $\cos \theta$ is the obliquity factor that is ignored most of the time -that is $\cos \theta \approx 1$ -. However this is a highly complex calculation when transformed to discrete domain -that is, the complexity is $\mathcal{O}(N^4)$ -, and hence the formulation is not practical to use in real-life applications. Therefore alternative approaches have emerged in the literature for faster holographic reconstruction.

2.3.1 Reconstruction using Fresnel approximation

Fresnel method is the most common way of performing hologram reconstructions in the literature. If the propagation distance d is large enough compared to the dimensions of the

object plane, then ρ can be approximated to d without loss of integrity. In the denominator of the division, ρ is directly substituted by d , and in the numerator the first two terms of the binomial expansion of ρ is used as follows:

$$\rho \approx d + \frac{(x - u)^2}{2d} + \frac{(y - v)^2}{2d}. \quad (2.15)$$

Ignoring the obliquity factor $\cos \theta$, the Rayleigh-Sommerfeld diffraction integral is simplified to a so-called Fresnel transform representation [31]:

$$I_d(u, v) = \frac{1}{i\lambda d} \exp(-i\frac{2\pi}{\lambda}d) \exp(-i\frac{\pi}{\lambda d}(u^2 + v^2)) \times \int \int I_0(x, y) \exp(-i\frac{\pi}{\lambda d}(x^2 + y^2)) \exp(-i\frac{2\pi}{\lambda d}(ux + vy)) dx dy. \quad (2.16)$$

When the constant terms in front of the integrals are further ignored, and two quadrature terms are defined as follows,

$$\begin{aligned} Q_i(u, v) &= \exp(-i\frac{\pi}{\lambda d}(u^2 + v^2)), \\ Q_o(x, y) &= \exp(-i\frac{\pi}{\lambda d}(x^2 + y^2)), \end{aligned} \quad (2.17)$$

the Fresnel transformation turns out to be a two dimensional Fourier transformation of $I_0(x, y)$ multiplied by a quadrature term in the phase:

$$I_d(u, v) = Q_i(u, v) \int \int I_0(x, y) Q_o(x, y) \exp(-i\frac{2\pi}{\lambda d}(ux + vy)) dx dy. \quad (2.18)$$

Once the problem is transformed into a Fourier transformation operation, then it is possible to utilize the Fast Fourier Transformation method to speed up the process. Fast Fourier Transformation method (FFT) is a fast way to evaluate discrete Fourier transformations, and it has $\mathcal{O}(N^2 \log N)$ complexity. Note that the equation 2.16 is still defined in the continuous domain and it should be converted to a discrete form to be able to perform numerical evaluation. Assuming the object plane has $N_x \times N_y$ points with Δx and Δy steps, and the camera plane is also sampled with the same number of points and with Δu and Δv steps, then the step-sizes of the two planes are related due to the Fourier transform relationship in the following way:

$$\Delta u = \frac{\lambda d}{N_x \Delta x}, \quad \Delta v = \frac{\lambda d}{N_y \Delta y}. \quad (2.19)$$

Introducing these substitutions in to the equation 2.16, the discrete form of the Fresnel equation is written as:

$$I_d(u, v) = Q_i(u, v) \sum_{x=-N_x/2}^{N_x/2-1} \sum_{y=-N_y/2}^{N_y/2-1} I_0(x, y) Q_o(x, y) \exp \left[i2\pi \left(\frac{ux}{N_x} + \frac{vy}{N_y} \right) \right], \quad (2.20)$$

where quadrature terms are defined as:

$$\begin{aligned} Q_i(u, v) &= \exp \left[-i\pi\lambda d \left(\frac{u^2}{N_x^2 \Delta x^2} + \frac{v^2}{N_y^2 \Delta y^2} \right) \right], \\ Q_o(x, y) &= \exp \left[-i\frac{\pi}{\lambda d} (x^2 \Delta x^2 + y^2 \Delta y^2) \right]. \end{aligned} \quad (2.21)$$

Moreover, the indices (x,y) in the summation in equation 2.20 are required to be handled since they should start from zero instead of $-\frac{N_x}{2}$ and $-\frac{N_y}{2}$. To correct this mismatch, a shift operation is performed on the indices, and the discrete Fresnel transformation is defined in the following way with the modified quadrature terms [51]:

$$\begin{aligned} Q'_i(u, v) &= \exp \left(i\pi \frac{N_x + N_y}{2} \right) \exp(-i\pi(u + v)) Q_i \left(u - \frac{N_x}{2}, v - \frac{N_y}{2} \right), \\ Q'_o(x, y) &= \exp(-i\pi(x + y)) Q_o \left(x - \frac{N_x}{2}, y - \frac{N_y}{2} \right), \\ I_d(u, v) &= Q'_i(u, v) \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} I_0(x, y) Q'_o(x, y) \exp \left[i2\pi \left(\frac{ux}{N_x} + \frac{vy}{N_y} \right) \right]. \end{aligned} \quad (2.22)$$

As indicated above, this transformation is in the form a discrete Fourier transformation multiplied with the quadrature term Q_o . However due to the sign of the multiplied phase term at the end, the operation is actually classified as an inverse Fourier transform. Hence finally as the last step, the inverse discrete Fourier transformation is substituted with the inverse FFT method, and the Fresnel transformation is then performed in $\mathcal{O}(N^2 \log N)$ with the following procedure:

$$I_d(u, v) = Q'_i(u, v) FFT^{-1} [I_0(x, y) Q'_o(x, y)]. \quad (2.23)$$

With this formulation, reconstruction using Fresnel approximation takes as much time as a single FFT operation and two point-wise multiplications with quadratic functions. However when a reconstruction is performed with the Fresnel method, a magnification depending on the distance d occurs on the scene, because the pixel dimensions Δu and Δv change by d . This may pose a significant problem for a variety of applications such

as continuously tracking particles in multiple depths, or for example, when performing autofocusing. Since each reconstruction with different depth will have different magnification level, a correction procedure might be essential when using this transformation in such applications. The transformation also has a minimum propagation distance limitation, that is, the Fresnel's approximation is only valid when the propagation distance is much larger than the dimensions of the object:

$$d^3 \gg \frac{\pi}{4}((x-u)^2 + (y-v)^2)^2. \quad (2.24)$$

2.3.2 Reconstruction using convolution

Convolution theorem can also be utilized for reconstructing holograms in a fast manner. The propagation of a light wave is considered to be a linear system in free space, and therefore it is possible to express the diffraction integral in the following way [51]:

$$I_d(u, v) = \int \int I_0(x, y) c_d(u, v, x, y) dx dy, \quad (2.25)$$

where $\cos \theta \approx 1$, and $c_d(u, v, x, y)$ is the impulse response function of the free-space for a given distance d :

$$c_d(u, v, x, y) = \frac{i}{\lambda} \frac{\exp \left[-i \frac{2\pi}{\lambda} \sqrt{d^2 + (x-u)^2 + (y-v)^2} \right]}{\sqrt{d^2 + (x-u)^2 + (y-v)^2}}. \quad (2.26)$$

Assuming the system is time-invariant, the diffraction integral can be treated as a convolution integral. However to perform numeric evaluations, the discrete version of the impulse response function must be used:

$$c_d(x, y) = \frac{i}{\lambda} \frac{\exp \left[-i \frac{2\pi}{\lambda} \sqrt{d^2 + (x - \frac{N_x}{2})^2 \Delta x^2 + (y - \frac{N_y}{2})^2 \Delta y^2} \right]}{\sqrt{d^2 + (x - \frac{N_x}{2})^2 \Delta x^2 + (y - \frac{N_y}{2})^2 \Delta y^2}}. \quad (2.27)$$

The propagation of the hologram can be computed directly by convolving $I_0(x, y)$ and $c_d(x, y)$, but it is not practical because this has a complexity of $\mathcal{O}(N^4)$. However according to the convolution theorem, the linear convolution of two signals is equal to the inverse Fourier transform of the multiplication of the two signals in frequency domain. Utilizing the FFT method for discrete Fourier transformations, the complexity of the convolution operation is reduced to $\mathcal{O}(N^2 \log N)$. The hologram reconstruction procedure is

then mathematically expressed as:

$$I_d(u, v) = FFT^{-1}[FFT(I_0(x, y)) \cdot FFT(c_d(x, y))]. \quad (2.28)$$

Hologram reconstruction with the convolution method takes as much time as three FFT operations and a point-wise multiplication with the impulse response function. Although not shown here, it is possible to define the impulse response function in frequency domain, and then the required number of FFT operations to perform a reconstruction would be reduced by 1. The main advantage of the convolution method is that the pixel dimensions Δu and Δv remain same for any reconstruction and therefore all of the reconstructions are of the same scale level. This is quite useful when comparing different reconstructions of the same hologram such as in autofocusing. Finally, the convolution method has a minimum propagation distance limitation due to the Nyquist sampling theory. To perform a valid reconstruction, distance d should adhere to the following restriction:

$$d \geq \max \left(\frac{N_x \Delta x^2}{\lambda}, \frac{N_y \Delta y^2}{\lambda} \right). \quad (2.29)$$

2.3.3 Reconstruction using angular spectrum

A third method for reconstructing holograms is the angular spectrum method. This method is known to numerically propagate waves more accurately than others when propagation distances are small. The method basically assumes the wave-field $I_0(x, y)$ is traveling on a plane along the positive z direction where it expresses the initial field as the angular spectrum of plane waves in x and y axes, and then it propagates those plane waves accordingly as the wave travels in z direction. The angular spectrum of the complex wave is calculated using the Fourier transform, where the transform variables are in terms of the direction cosines of the wave vector [31]:

$$A_0\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) = \int \int I_0(x, y) \exp \left[-i2\pi \left(\frac{\alpha}{\lambda} x + \frac{\beta}{\lambda} y \right) \right] dx dy, \quad (2.30)$$

where A_0 is called the angular spectrum of the wave I_0 , I_0 is the wave at $z = 0$ plane, λ is the wavelength, and α and β are the direction cosines of the wave vector in x and y axes respectively. The propagation along the z -axis is performed by shifting the phases of the

plane waves in the angular spectrum of I_0 :

$$A_d\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) = A_0\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) \exp\left(i\frac{2\pi}{\lambda}d\gamma\right), \quad (2.31)$$

where γ is the direction cosine in z axis. The three direction cosines of the complex wave are interrelated through:

$$\alpha^2 + \beta^2 + \gamma^2 = 1. \quad (2.32)$$

Let $m = \alpha/\lambda$ and $n = \beta/\lambda$, then the direction cosine in z axes can be substituted by:

$$\gamma = \sqrt{1 - (\lambda m)^2 - (\lambda n)^2}. \quad (2.33)$$

The propagation operation is then expressed as:

$$A_d(m, n) = A_0(m, n) \exp\left(i\frac{2\pi}{\lambda}d\sqrt{1 - (\lambda m)^2 - (\lambda n)^2}\right). \quad (2.34)$$

Finally, it is required to back-transform the propagated angular spectrum to the spatial domain using inverse Fourier transformation, that is:

$$I_d(u, v) = \int \int A_d(m, n) \exp[i2\pi(mu + nv)] dm dn. \quad (2.35)$$

Note that the integral will be evaluated only for the region $\alpha^2 + \beta^2 < 1$, which means that the wave is a propagating one. The regions where this condition is not met express evanescent waves, but this is not related to our topic. To sum up the overall reconstruction procedure, wave propagation using the angular spectrum can be described by the following [45]:

$$I_d(u, v) = FFT^{-1}[FFT(I_0(x, y)) \cdot T_d(m, n)], \quad (2.36)$$

where $T_d(m, n)$ is the wave propagation function of the angular spectrum method, that is:

$$T_d(m, n) = \begin{cases} \exp\left(i\frac{2\pi}{\lambda}d\sqrt{1 - (\lambda m)^2 - (\lambda n)^2}\right), & \sqrt{m^2 + n^2} < \frac{1}{\lambda} \\ 0, & otherwise. \end{cases} \quad (2.37)$$

Reconstruction using the angular spectrum method consumes as much time as two FFT operations and a point-wise multiplication with a transfer function. The method works more accurately for small propagation distances and there is no minimum distance limitation, which is a superior feature compared to the other reconstruction methods described

above considering that there is a minimum propagation distance restriction for them. The angular spectrum method also preserves the axial dimensions of the hologram similar like the convolution method. These two features make the angular spectrum method highly attractive in microscopic applications as well as in autofocusing. For these reasons, hologram reconstructions are performed by the angular spectrum method in the rest of this thesis.

2.4 Computer Generated Holograms

A computer generated hologram do not contain physical data, rather, the hologram is synthetically produced in computer environment by calculating the interference pattern of two complex waves. While one of the waves is set as a -traditionally simple- reference wave, the other wave contains the object data and it can be any digital image in the intensity and phase parts. Computer generated holograms are often used for simulation purposes and they are also referred multiple times in this thesis. Here, the computer generated hologram process is described briefly.

Suppose that we want to synthesize a computer generated hologram of an image using a uniform planar reference wave. First, a digital gray-scale image is converted to a complex representation and its phase part is assumed to be uniform. Then this complex image is propagated to the desired distance using equation 2.36. After that, an interference pattern is generated by multiplying the diffracted wave with the reference wave that is defined in discrete domain with the following formulation:

$$R(m, n) = \exp \left[-i \frac{2\pi}{\lambda} [\sin(\alpha)m\Delta m + \sin(\beta)n\Delta n] \right], \quad (2.38)$$

where λ is the wavelength, Δm and Δn are the pixel dimensions, and α and β are the angles of the reference wave between x-z and y-z axes respectively. For an in-line hologram, α and β are equal to 0.

Before reconstructing the computer generated hologram, filtering is needed to be performed. It is quite easy to apply the phase-shifting holography technique to computer generated holograms. Basically, three holograms with reference wave phases 0, $\pi/2$ and π are created. The offset in the phase of the reference wave defined in equation 2.38 is

assumed to be 0, and $R(m, n)$ is multiplied with $e^{i\frac{\pi}{2}}$ and $e^{i\pi}$ values to generate the phase shifted reference waves. After the equation 2.10 is applied, the resultant filtered hologram is ready for reconstruction.

To reconstruct the hologram, the filtered hologram is simply multiplied with $R(m, n)$ for simulating illumination, and then it is back-propagated to the camera plane using again the equation 2.36. Typically, if the hologram is generated k cm away, now it is required to be reconstructed at $-k$ cm. It is essential to know the amount of back-propagation in any holographic reconstruction, because any propagation distance other than the required, yields a useless unfocused image.

Chapter 3

DIGITAL HOLOGRAPHIC MICROSCOPE

Digital holographic microscopy uses the same principles of digital holography. The setup of a DHM is almost identical to a digital holography setup. However, one important difference is that the object wave is magnified by a pin-hole or a microscope lens. Moreover, the propagation distances used in holographic reconstruction is very small, and therefore the angular spectrum method is best suited for applications in digital holographic microscopy.

In the first section of this chapter, optical setup of the digital holographic microscope (DHM) is described in detail. In the second section, a commodity graphics card is integrated to DHM system for performing fast calculations. The developed interface for controlling the DHM is presented in the third section. Finally in the fourth section, some reconstruction experiments on holograms recorded by the DHM are shown.

3.1 Optical Setup

We have realized a Mach-Zehnder interferometer whose schematic is shown in figure 3.1. The components of the microscope are: two beam-splitters, two microscope objectives, two mirrors, a phase-modulator, a high-voltage amplifier, a National Instruments USB-6251 (NI), a sample holder and a digital camera. The laser beam is divided into two arms by a beam-splitter (BS1), where the reference wave is modulated by a phase modulator

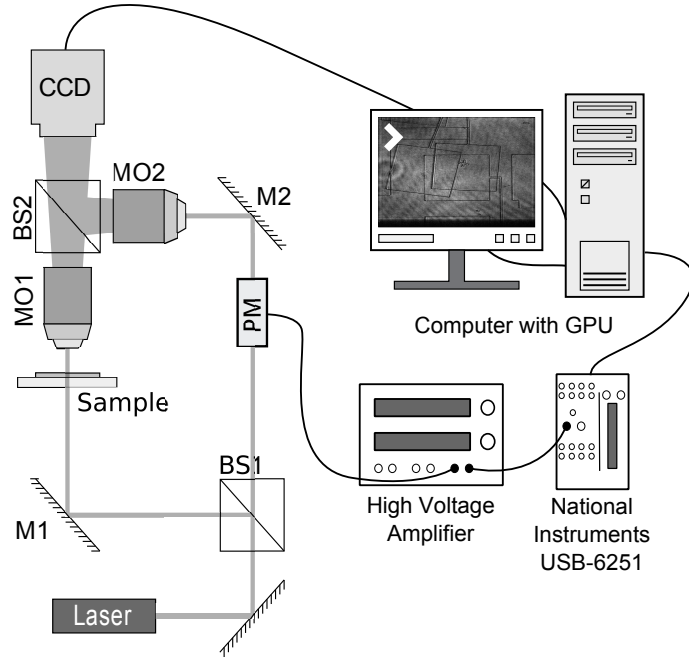


Figure 3.1: A Mach-Zehnder interferometer based digital holographic microscope setup that can perform phase-shifting holography. The laser beam is divided into two arms by a beam-splitter (BS1). The object wave is transmitted through the specimen (S) while the phase of the reference wave is modulated by a phase-modulator (PM). The phase-modulator is connected to a high-voltage amplifier which is connected to National Instruments USB-6251 that is controlled by a personal computer. The beams are magnified by identical microscope objectives and they are superimposed on each other using another beam-splitter (BS2). The formed interference pattern is recorded by a CCD camera that is connected to the personal computer.

(PM) and the object wave is transmitted through the specimen. The beams are magnified by microscope objectives before superimposed on each other by the second beam-splitter (BS2). Finally, the interference pattern of the wavefields is recorded by a CCD camera.

The microscope is transmission based and it is able to record holograms of transparent objects and biological cells without staining. A He-Ne laser with linear polarization is used for the illumination where the wavelength is 632.8 nm. To be able to perform phase-shifting holography, an optoelectronic phase-modulator is placed on the path of the reference beam, and the phase is modulated by applying high voltages on it. This is achieved via an electronic equipment -National Instrument USB 6251 (NI)- that is connected to the computer by a USB cable. An analog port of the NI is also connected to a 40x high-voltage amplifier and the optoelectronic crystal is connected to the voltage

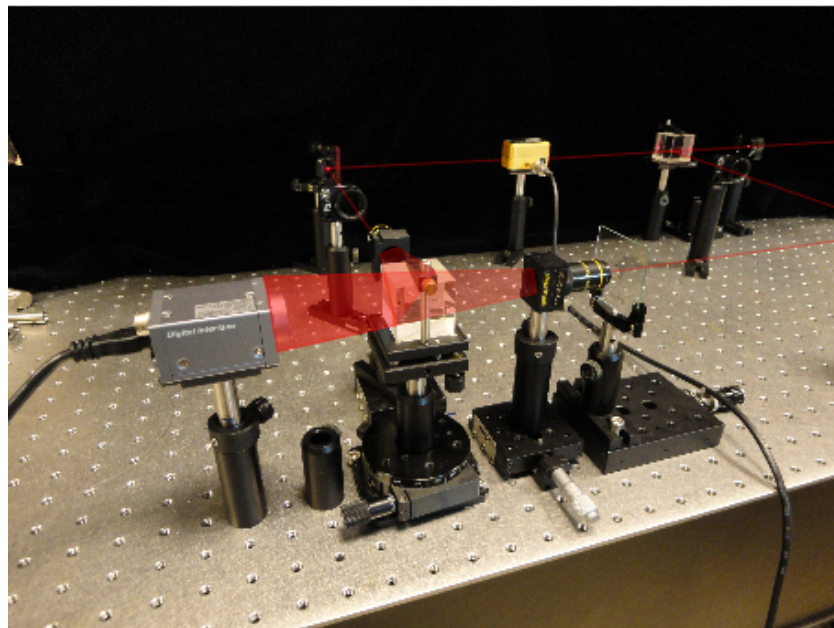
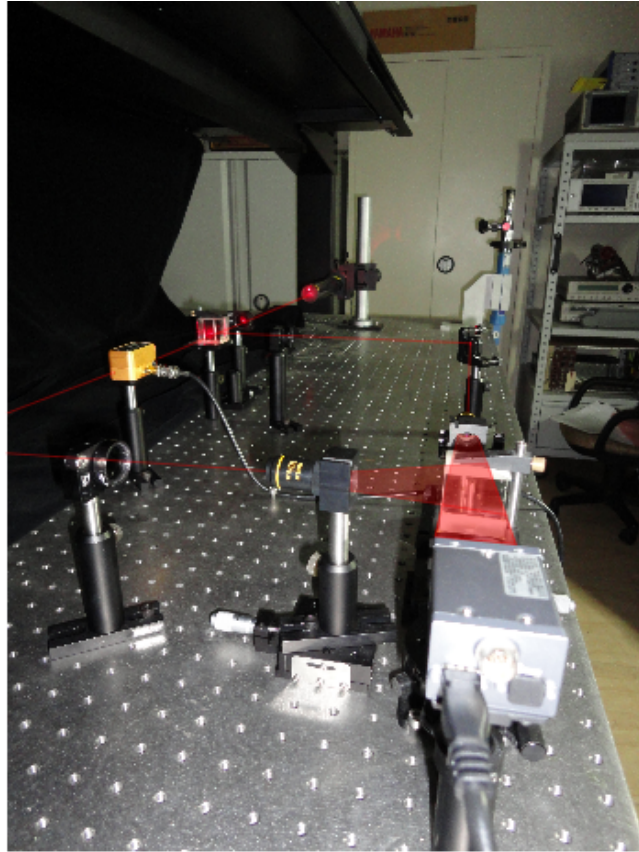


Figure 3.2: Digital holographic microscope. The laser beam is animated for better illustration.

amplifier. To perform phase-shifting, a desired voltage is set on the control interface of the software, and that value is digitally transferred to NI. Then, NI feeds that voltage to the high voltage amplifier, and the high voltage amplifier drives the phase-modulator. The

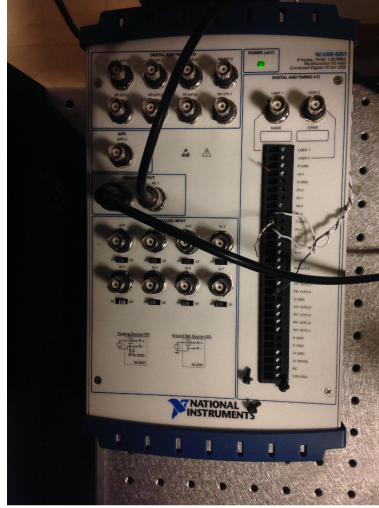


Figure 3.3: National Instruments USB 6251 (NI) is used for driving the optoelectronic phase modulator. The device is connected to the personal computer via USB cable. After the desired voltage information is fed from the computer digitally, NI applies that much voltage from its analog port A01 to a high-voltage amplifier which is connected to the phase modulator.

opto-electronic modulator can operate at up to 50 MHz rate. The required voltages for exact π and $\frac{\pi}{2}$ phase-shifts are calibrated by-hand on the software.

The object beam in the microscope is magnified by a 10x microscope objective (MO1) with a numerical aperture of 0.25 for recording holograms of small particles. However,

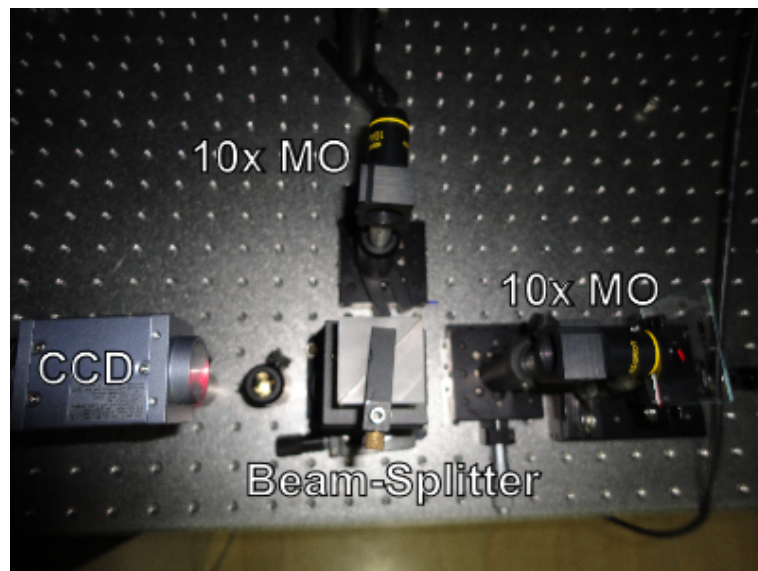


Figure 3.4: Two identical microscope objectives (MO) are placed on the path of the object and the reference beam for equalizing the wavefront curvatures of the waves. The microscope objectives are symmetrical with respect to the beam-splitter in x, y and z axes.

this results a deformed and circular interference pattern because the curvatures of the reference and the object beams do not match. Therefore, another identical microscope objective (MO2) is placed on the path of the reference beam as well. The two microscope objectives are located precisely symmetrical in x, y and z axes, interference pattern is formed as a planar one.

A Sony S800 camera is used for digitally capturing the interference patterns. This is a monochromatic camera that has 1024×1024 pixels with pixel dimensions of $3.75 \mu\text{m} \times 3.75 \mu\text{m}$. According to this configuration, the microscope has $2.2 \mu\text{m}$ lateral resolution in both axes, and the field of view is $320 \mu\text{m} \times 320 \mu\text{m}$ with overall magnification factor of

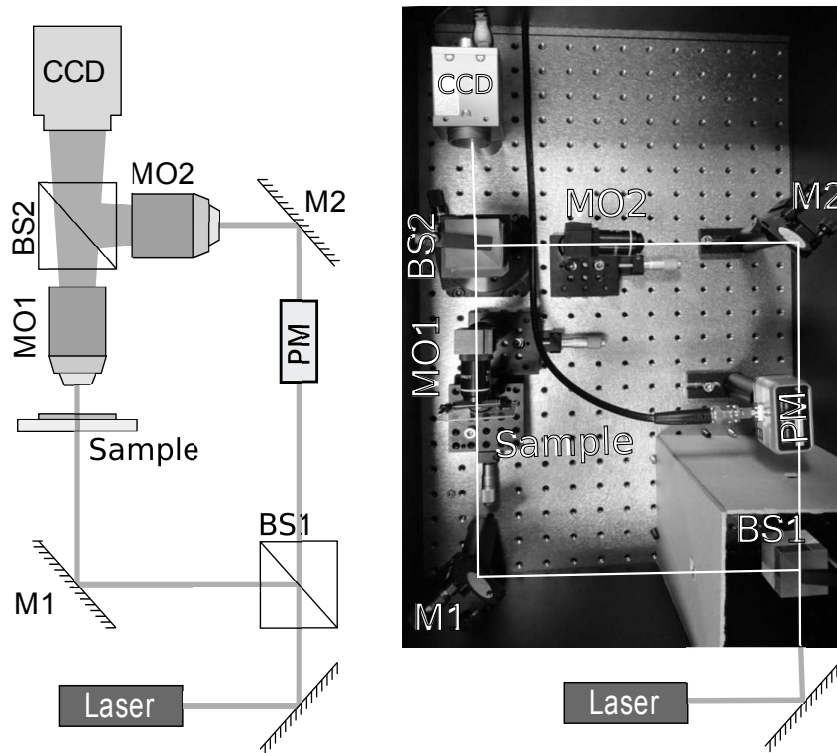


Figure 3.5: The optical components of the transmission based digital holographic microscope (DHM) setup are shown in correlation with the schematic. The laser beam is divided into two arms by a beam-splitter (BS1). The reference wave is modulated by a phase modulator (PM) and the object wave is transmitted through the specimen. Before joining the two arms of the interferometer by a second beam-splitter (BS2), object image is magnified by the microscope objective (MO1). In order to compensate the wavefront curvatures, a similar objective (MO2) is placed on the reference beam path. Finally, interference of the two beams is recorded by a CCD camera that is connected to a personal computer.

twelve. The camera can capture a maximum of 15 frames per second (maximum exposure time of 66 ms). Since three holograms are recorded for phase-shifting holography, a maximum of 5 filtered holograms can be acquired per second.

In overall, a Mach-Zehnder interferometer based digital holographic microscope is built, and it operates in transmission mode. The exact specifications of the microscope are presented in table 3.1. Moreover, the optical components of the microscope are shown in Fig. 3.5 for better correlation with the illustration. However, this optical setup is highly sensitive to disturbances in the air and the ground. Even small vibrations are captured by the interferometer which results in the failure of capturing precise phase-shifted holograms. To isolate the vibrations from the ground, the setup is built on an optical table that is stabilized by compressed air. Moreover to compensate the vibrations in the air, the setup is enclosed in a wooden box. The box is painted to black such that the reflection of the light within the system is minimized. Photographs of the final system is presented in figure 3.6 and figure 3.7.

Table 3.1: Specifications of the digital holographic microscope.

DHM feature	Specification
Mode	Transmission
Wavelength	632.8 nm
Filtering	Phase-shifting
Capture rate	5 holograms per second
Camera pixel dimensions	$3.75\mu\text{m} \times 3.75\mu\text{m}$
Pixel resolution	1024×1024
Lateral resolution	$2.2 \mu\text{m}$
Field of view	$320 \mu\text{m} \times 320\mu\text{m}$
Magnification	≈ 12

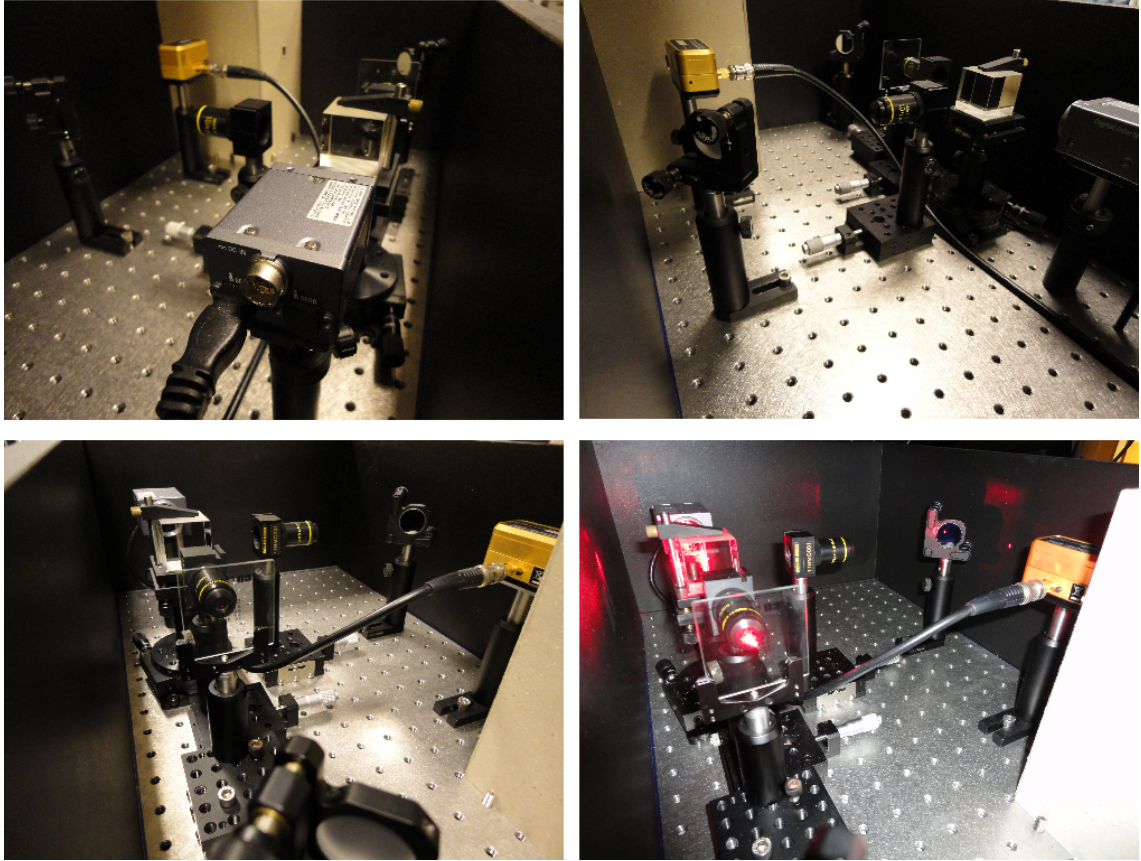


Figure 3.6: Snapshots of the digital holographic microscope.

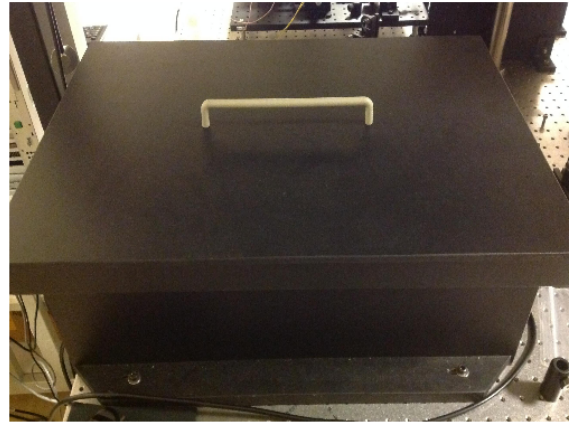
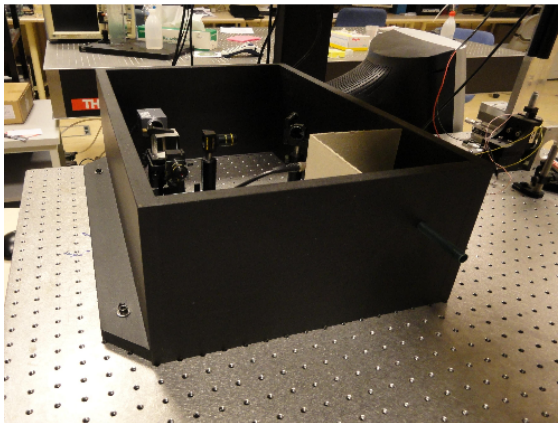
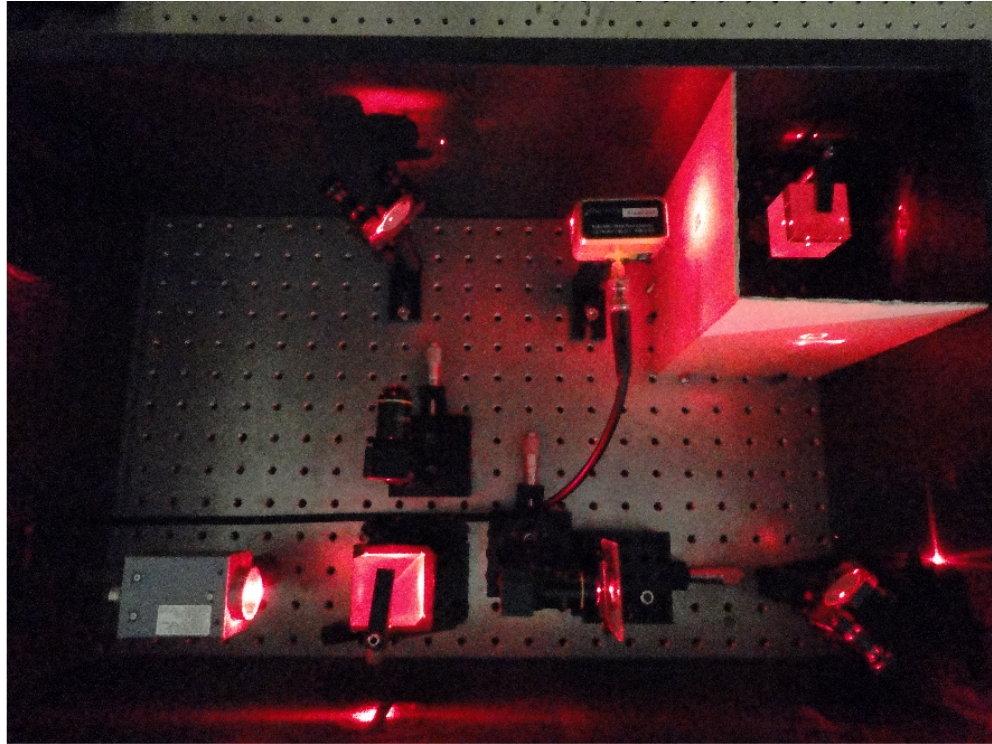


Figure 3.7: The digital holographic microscope is enclosed within a black box for reducing the effects of vibrations in the air.

3.2 Integration of the Graphics Card to the DHM

The digital holographic microscope we built, utilizes the graphics card of the computer for performing hologram reconstructions in fast manner. Each graphics card has a graphics processing unit (GPU) that contains hundreds of processors and a high bandwidth memory. Utilizing the parallel computation architecture, GPUs exceed the performance of CPUs by orders of magnitude while performing computations on large chunks of data. One important consideration is that the desired algorithm should be able to take the form of a parallel code structure to utilize the resources available in the graphics processing unit. Hologram filtering, for example, is a point-wise operation and it is perfectly suited for parallelization since the operation of the processors are not dependent on each others outcome. Hologram reconstruction, on the other hand, is a more complex operation that involves fast Fourier transformations (FFT). Nvidia provides a platform called Compute Unified Device Architecture (CUDA) for developing GPU-centric softwares on graphics cards, and the Nvidia implementation of the fast Fourier transformation (CUFFT) has impressive performance. Therefore, CUFFT library is used for performing reconstructions

```
static __global__ void fill_angular_response( float2 *in, int N, float distance, float lambda, float del_x, float del_y ) {
    int i = threadIdx.x + blockIdx.y*(thread_size*thread_size);
    int j = blockIdx.x;
    int index = i + j*N; // The position on the transfer matrix is calculated using the block and thread IDs.

    float temp;
    float fx, fy;
    if ( i < N && j < N ) {
        fx = ( i - N/2 ) / ( del_x * N );
        fy = ( j - N/2 ) / ( del_y * N );
        if ( ( 1 / ( lambda * lambda ) - lambda * lambda * (fx * fx + fy * fy)) > 0 ) {
            // Transfer function is calculated
            // as T = temp = 1 x exp ( i x alpha )
            temp = ( 2.0f * PI * d / lambda ) * sqrt(1 - ( fx * fx + fy * fy ) * ( lambda * lambda ) );

            // The calculated angular value is converted from phasor domain to cartesian coordinates
            float2 current = make_float2( cosf(temp), sinf(temp) );
            in[index] = current;
        } else {
            float2 current = make_float2(0,0);
            in[index] = current;
        }
    }
}
```

Figure 3.8: The calculation of the transfer function of the angular spectrum method implemented in CUDA language for execution in GPU (Nvidia GTX 660 Ti).

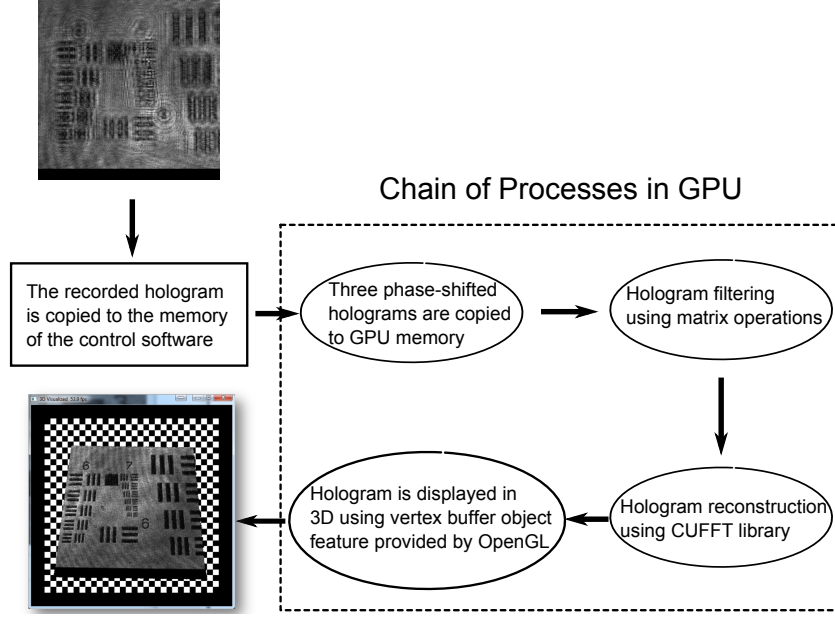


Figure 3.9: The reconstruction procedure of a recorded hologram using the graphics processing unit.

of holograms recorded by the DHM.

The parallelization of the hologram reconstruction algorithm is organized by a user-specified-kernel function which is called by the main CPU of the computer. The kernel functions are written in CUDA programming language, and then the very same kernel function is executed throughout the processors of the GPU. Processors inside the GPU are organized according to user-specification in $m \times n$ blocks, and each processor executes t number of concurrently processed threads [52]. With this organization, a virtually 3D computing architecture with dimensions of $m \times n \times t$ is formed where each cell executes the same kernel function. The computational diversity between the cells are achieved by utilizing the index information of the cell as a parameter of the kernel function. As an example, calculation of the transfer function of the angular spectrum method is presented in CUDA language in figure 3.8.

The flow of hologram data is organized as follows. When a hologram is recorded by the camera, it is immediately copied to the memory of the graphics card. Hologram filtering and reconstruction process starts on GPU after three holograms are shot by the camera, and transferred to the GPU memory. After three holograms are copied, the hologram filtering procedure is executed using matrix summation and difference operations. After

that, hologram is reconstructed by one of the three methods described earlier, where FFTs are calculated by the CUFFT library. The reconstruction is not copied back to the memory of the control interface for displaying purposes thanks to the CUDA and OpenGL interoperability services. Instead, the control software directly reads the reconstruction from the GPU for displaying purposes utilizing the Vertex Buffer Object (VBO) feature. The program flow can be described as in figure 3.9.

3.3 Control Interface of the DHM

The DHM is controlled with a software developed in C# programming language. A snapshot of the is presented in figure 3.10. On the upper part of the program, the real-time feed of the camera is shown. After the appropriate voltages are set on the lower-left part, the phase-modulator is automatically modulated by the software in synchronization with the camera feed. After three phase-shifted holograms are shot by the camera, the software immediately filters the hologram, reconstructs it, and displays the result. All of the three hologram reconstruction methods (the Fresnel approximation, the convolution and the angular spectrum methods) are implemented. The propagation distance for the reconstruction is set by the bar located in the middle of the software window, and real-time feed is possible since the software utilizes the graphics engine (Nvidia GTX 660 Ti) for computing the hologram reconstructions. As presented in figure 3.11, the calculated object wave is displayed in 3D using Vertex Buffer Objects (VBO) provided by the OpenGL library, where the magnitude of the object wave is set as the color map, and the phase of the object wave is set as the height map.

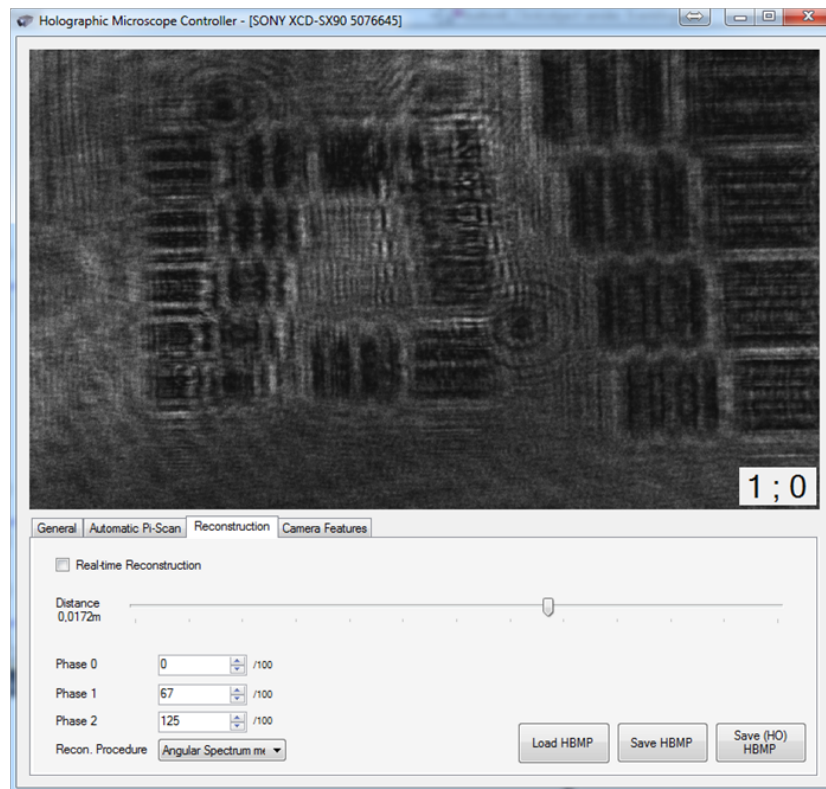


Figure 3.10: A snapshot of the controlling interface of the software developed for operation of the DHM.

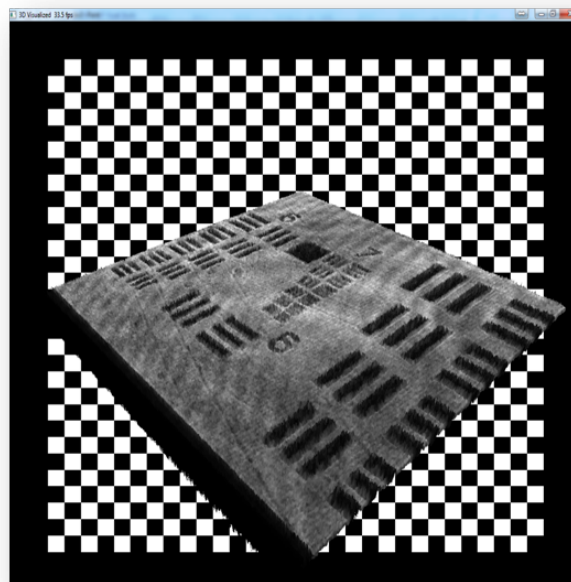


Figure 3.11: A snapshot of the display window of the software developed for the DHM.

3.4 Hologram Reconstruction Experiments with the DHM

In this section, the three reconstruction methods are applied to the holograms recorded by the digital holographic microscope. In figure 3.12, the holographic image of onion cells is presented, and in figure 3.13, the holographic image of USAF resolution chart is shown. Finally in figure 3.14, another holographic image of USAF resolution chart is presented where the Fresnel approximation and the convolution methods are invalid because the minimum propagation distance requirement is not met. The minimum propagation distance restriction for the convolution method is calculated as 2.28 cm by equation 2.29 for this DHM configuration, where:

$$\begin{aligned}\Delta x &= \Delta y = 3.75 \times 10^{-6}, \\ N_x &= N_y = 1024, \\ \lambda &= 632.8 \times 10^{-9}, \\ d_{min} &= 0.0228 \text{ m}.\end{aligned}\tag{3.1}$$

The minimum distance limitation for the Fresnel approximation is also around 1.5 cm for our configuration. However, the angular spectrum method does not pose a limitation on

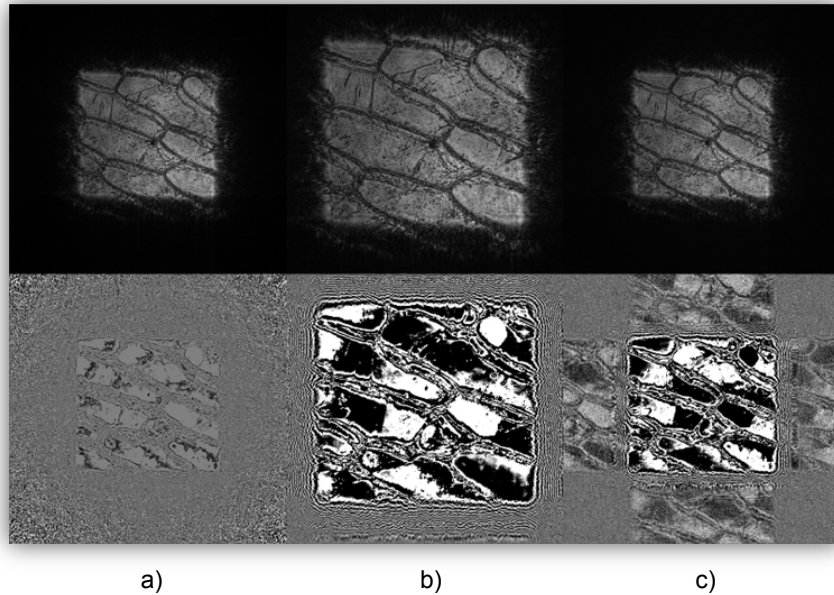


Figure 3.12: A hologram of onion cells is recorded 3 cm away from the recording plane. The magnitude (top row) and the phase (bottom row) of the holographic reconstructions by the a) Angular spectrum, b) Fresnel approximation and c) Convolution methods are presented.

the propagation distance, and it still works in this example. Therefore, we can say that angular spectrum method is better suited for microscopic applications.

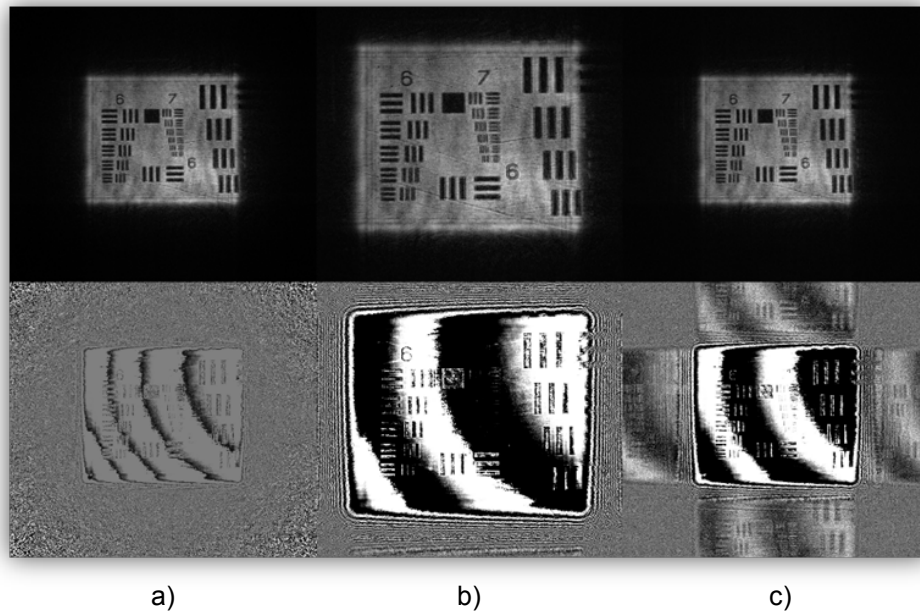


Figure 3.13: A hologram of USAF resolution chart is recorded 4.5 cm away from the recording plane. The magnitude (top row) and the phase (bottom row) of the holographic reconstructions by the a) Angular spectrum, b) Fresnel approximation and c) Convolution methods are presented.

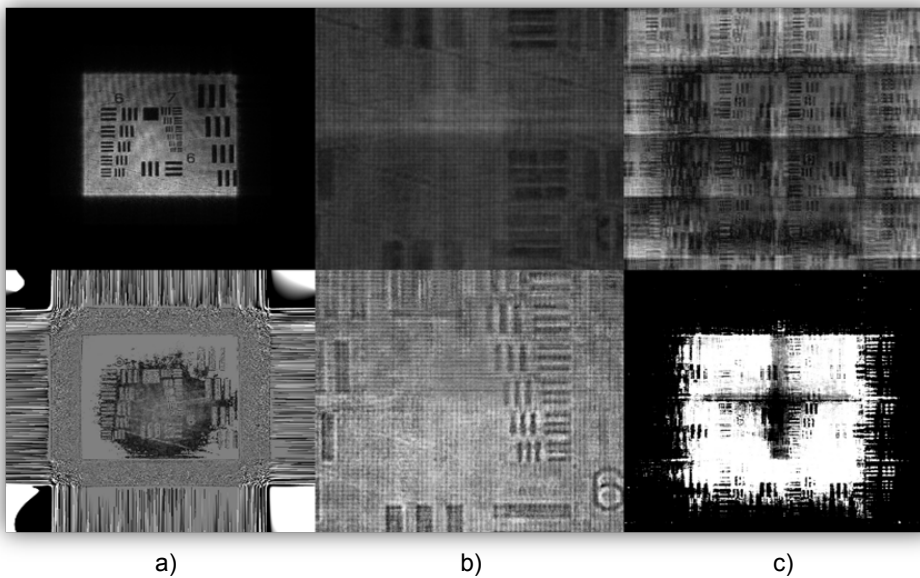


Figure 3.14: A hologram of USAF resolution chart is recorded 1 cm away from the recording plane. The magnitude (top row) and the phase (bottom row) of the holographic reconstructions by the a) Angular spectrum, b) Fresnel approximation and c) Convolution methods are presented.

Chapter 4

AUTOFOCUSING

Focusing is an important problem for any imaging application that is going to be perceived by a human, and it is a long-standing topic in image processing literature. Technically, focusing can be defined as adjusting the position of the lenses according to the scene such that the captured image is perceived with the most quality. For a human eye, quality of an image is generally associated with its sharpness and contrast. In 1974, Muller and Buffington [53] proposed using sharpness metrics for measuring the in-focusness of images. According to their claim, when a sharpness metric is used to evaluate various captures of a scene, the metric should produce a global extrema only for the focused image. Although there have been excessive development on this topic today, the underlying idea has remained the same for focusing. Today in photography and microscopy, the position of the lenses or the stage are adjusted successively as the sharpness of acquired images are evaluated by a sharpness metric. To this extent, image processing literature contains numerous metrics and methods for focusing intensity-only images [54, 36, 37].

In this chapter, the most common sharpness metrics in literature are investigated and application of those metrics on digital holograms is explained. It is experimentally shown that autofocusing of holograms is possible using traditional sharpness metrics.

4.1 Focusing of Digital Holograms

In digital holography, focusing problem is simply reduced to choosing the most appropriate reconstruction distance for a hologram. If this focus distance is known priori, then

a focused image is achieved by a single reconstruction. Otherwise, it can be estimated manually by trying out different reconstruction distances by hand. A clear advantage of holographic imaging is that only one capture of a scene is sufficient for finding this 'most appropriate reconstruction distance', since the hologram can be numerically reconstructed at any arbitrary distance as if it was shot from a different distance. Thus, it is possible to compare the in-focusness of various reconstructions of a single recorded hologram by applying Muller and Buffington's analogy. This means that the focusing procedure of a hologram can be easily automated, and holograms can be focused without knowing the true-focus-distance priori.

Holographic imaging literature contains examples of applying traditional sharpness metrics to holograms for focusing purposes. Langehanenberg [55] and Memmolo [56] have chosen the normalized-variance as the most reliable traditional metric for autofocus-ing in digital holography. In another work, a self-entropy metric is used by Gillespie and King [38] to focus digital holograms. There are also complex wavelet-transform-based methods developed for focusing digital holograms [39, 57] as well as frequency spectrum based metrics [55]. All of these methods take advantage of the sharp edges and high contrast of focused images, therefore it is possible to apply these techniques locally when there are multiple objects in the scene. There is also another focusing method proposed by Dubois [40] for focusing holograms. Dubois' method analyzes the integral power of the hologram as a focus measure, and hence it cannot be applied locally for a specific part of the hologram.

4.2 Autofocusing in Digital Holography

To obtain a quality image of an object, a hologram must be reconstructed at a specific distance -that is the optical distance from the object plane to the camera plane-. This distance must be correct on the order of sub-millimeters, otherwise the reconstructed image appears blurry and out of focus. However if the focus distance of a hologram of an object is not known a priori, it can be found by estimating its sharpness curve. As illustrated in figure 4.1, a hologram is reconstructed at several candidate distances over a depth in-

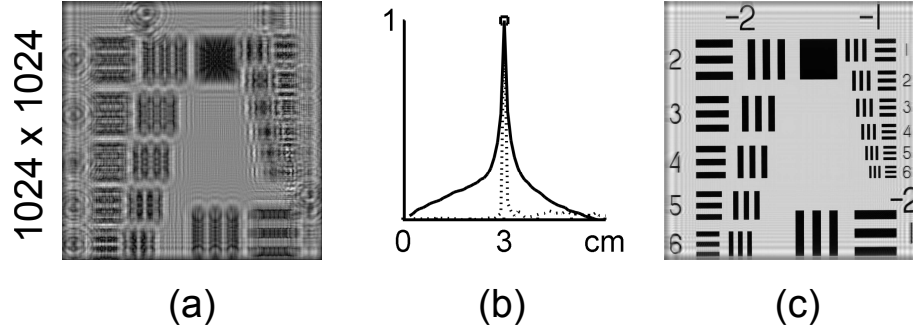


Figure 4.1: Autofocusing is shown on a computer generated hologram of the USAF chart. The hologram is reconstructed between the interval 0 to 6 cm with 0.1 mm step-size. Then each reconstruction is evaluated by the normalized-variance metric and the sharpness curve is formed. Finally, the hologram is reconstructed at the propagation distance that corresponds to the peak of the sharpness curve.

interval, and then the focus quality of each reconstruction is quantitatively evaluated using a sharpness estimation function. The sharpness curve is then formed by plotting those sharpness values against the reconstruction distances. Finally the corresponding reconstruction distance on the sharpness curve which yields the best quantitative result is depicted as the true-focus-distance (TFD) of the hologram. Typically, sharpness estimation functions approach to a global extrema as the reconstruction distance approaches to the true-focus-distance (TFD) of a hologram.

An important consideration about focusing of holograms is that, a hologram contains both the intensity and the phase information about a scene whereas the traditional sharpness metrics in literature have been developed to work on intensity-only images. In this work, only the magnitude of the reconstructed images are used for assessing the sharpness. Focusing using the phase of the reconstructed images is also possible, but then the process gets more complex since the phase image contains phase-wraps. In that case phase-unwrapping has to be performed on each reconstruction before estimating its sharpness. Otherwise, those phase-wraps are interpreted erroneously as features in the image and the sharpness metrics produce unreliable results. Not only phase-unwrapping is a very time consuming procedure, it can also be unreliable depending on the phase profile and therefore it is better to be avoided for practical reasons. Another consideration about focusing holograms is that, holograms may contain images of phase objects instead of

opaque objects. Yet in those cases, the intensity images of the reconstructions are still useful for finding the true-focus-distance, because then the sharpness metrics take advantage of the invisibility of phase objects in the intensity image. For a hologram of a transparent object, the intensity-part of its reconstruction yields the least image features when the reconstruction is at focus. Therefore generally for phase objects, the true-focus-distance is identified by a global minima instead of a global maxima. For these reasons, it is highly attractive to use only the magnitude-part of the reconstructed images for focusing motives.

4.2.1 Sharpness Metrics

Eleven different numerical methods for computing a scalar value are explained here for evaluating the quality of an image [41]. These functions are grouped under derivative based, histogram based, power base and statistics based families according to their approach to the problem. In addition to those families, there is also the wavelet-transform-based focus metric family. However wavelet-based methods are computationally much more complex and they require a priori work on the subject as opposed to the traditional metrics that apply a relatively simple operation on the image. Therefore wavelet-transform-based methods are not included in this analysis.

Derivative-based metrics

The main idea of derivative-based metrics is that the in-focus images will contain strong edges and the local gradient of a focused image will be higher than its lesser-focused versions. Typically, the energy of the first or the second order derivatives is expected to increase since the high-frequency content in the image is lost as it gets blurred. Derivative-based metrics are susceptible to noise more than the other metrics, and thus variations of the first-order gradient function are proposed in literature for compensating the susceptibility to noise. In this work, gradient-squared, Brenner gradient, Tenenbaum gradient and energy of Laplacian metrics are investigated as described in Tab. 4.1.

Table 4.1: The description and the mathematical representation of the derivative based focus metrics used in this work are listed, where $O_{x,y}$ indicates the magnitude of the object image at pixel location (x, y) .

<i>Gradient-squared</i>	The function sums up the values that are greater than some threshold which is set proportional to the maximum squared-value in the gradient image.	$M_{\nabla} = \sum_x \sum_y [O_{x+1,y} - O_{x,y}]^2$
<i>Brenner gradient</i>	This function looks for wide edges in the image and it is more robust to noise than gradient function [58].	$M_B = \sum_x \sum_y [O_{x+2,y} - O_{x,y}]^2$
<i>Tenenbaum gradient</i>	Sobel operators compute the gradient of an image in X and Y axes, and they look for both wide and long edges in the image. Here, the energy of the image is computed after Sobel operators are applied [59].	$M_T = \sum_x \sum_y [O_{x,y} * S_X]^2 + [O_{x,y} * S_Y]^2$ $S_X = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, S_Y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$
<i>Energy of Laplacian</i>	This function computes the energy of the second-order derivative of the image. The discrete approximation of the second-order derivative is obtained by convolving the image with the given Laplacian mask.	$M_L = \sum_x \sum_y [O_{x,y} * L]^2, L = \begin{bmatrix} -1 & -4 & -1 \\ -4 & 20 & -4 \\ -1 & -4 & -1 \end{bmatrix}$

Histogram-based metrics

The histogram of images can also be used for assessing their sharpness [35]. The histogram based metrics basically compute the histogram of an image with k histogram bins ($O_{x,y} \rightarrow H(k)$), and then a simple function such as the range of bins or the entropy of the distribution of gray-levels is calculated on the histogram (see Tab. 4.2). The histogram of focused images is expected to contain a broad range of gray-levels, and the distribution of the gray-levels is expected to be more balanced compared to out-of-focus images. For example, as an image goes out of focus, the image will converge to a uniform gray-level value and all its pixels will eventually fit into a single histogram bin. The effect of noise on the histogram-based metrics is a function of the total number of histogram bins chosen. They are inversely proportional such that the effect of noise gets lesser as the number of histogram bins is chosen smaller, but this is at the expense of more high-frequency information. Generally speaking for gray-scale images, it is safe setting k to 64 for gray-scale images, since the human-eye is responsive to approximately 64 different gray-levels.

Table 4.2: The description and the mathematical representation of the histogram based focus metrics used in this work are listed, where $O_{x,y}$ indicates the magnitude of the object image at pixel location (x, y) .

<i>Histogram range</i>	The function computes the range between the first and the last non-zero histogram bins [35].	$M_{HR} = \max_i (i H(i) > 0) - \min_i (i H(i) > 0)$
<i>Histogram entropy</i>	This function computes the entropy of the distribution of gray-levels in an image. It is assumed that different gray-levels occur with more balanced odds and entropy is the lowest when the image is on focus [35]. This assumption is generally true except for some cases where the focused image is composed of just a few histogram bins. The metric is useful in those cases as well, because then the out-of-focus images of the object have lesser entropy and the measure is maximized when the image is focused.	$M_{HE} = - \sum_k p_k \log(p_k),$ $p_k = \frac{H(k)}{N_x N_y},$ <p>where p_k is the probability of randomly picking a pixel with the gray level of k^{th} bin.</p>

Power-based metrics

This class of functions are the simplest metrics to calculate, because they use the power of an image as a focusness measure instead of seeking features. Integral power and thresholded pixel count metrics are grouped under this family, and their mathematical representations are given in Tab. 4.3.

Table 4.3: The description and the mathematical representation of the power based focus metrics used in this work are listed, where $O_{x,y}$ indicates the magnitude of the object image at pixel location (x, y) .

<i>Integral power</i>	The integral power of a hologram reaches a global minimum for amplitude objects and a global maximum for phase objects when the object is focused [40].	$M_P = \sum_x \sum_y O_{x,y}$
<i>Thresholded pixel-count</i>	The number of pixels which have greater or lesser intensity than a given threshold is counted as the quality measure [36]. Depending on the choice of threshold direction, the metric is expected to produce either a global maximum or a global minimum at the true focus distance.	$M_{PC} = \sum_x \sum_y \tau[O_{x,y}]$ $\tau[O_{x,y}] = \begin{cases} 1, & O_{x,y} \leq \theta \\ 0, & otherwise \end{cases}$

Statistics-based metrics

Statistical metrics use the mean gray-level value of an image for comparing the in-focusness. As an image goes out of focus, it converges to a single gray-level, and thus the variance and contrast are minimized. Correlation of an arbitrary pixel with its neighbor is also lesser in an out-of-focus image. On the other hand, focused image possess greater contrast and neighboring pixels are correlated more with each other in sharp images. The statistical sharpness metrics are robust to effects of noise compared to other metrics. Normalized-variance, deviation based correlation and auto correlation functions listed in Tab. 4.4 belong to this family.

Table 4.4: The description an the mathematical representation of the statistics based focus metrics used in this work are listed, where $O_{x,y}$ indicates the magnitude of the object image at pixel location (x, y) .

<i>Normalized-variance</i>	This function calculates the variance of the image and divides it by the mean intensity value μ . This division compensates the different average intensity levels between different images and favors low mean intensities while emphasizing high variance [36, 35].	$M_{NV} = \frac{1}{\mu} \sum_x \sum_y [O_{x,y} - \mu]^2$
<i>Deviation based correlation</i>	Correlation of adjacent pixels are compared to the correlation of a constant intensity image that has gray level value of mean intensity μ at all pixels [60]. The function produces greater values for images with higher variance.	$M_{DBC} = \sum_x \sum_y O_{x,y} \cdot O_{x+1,y+1} - \mu\mu $
<i>Auto correlation</i>	This function computes the difference between the correlation of a pixel with itself and its neighbor [60]. The difference is expected to increase as the image comes to focus and gains in contrast. In essence, this operation corresponds to the multiplication of $O_{x,y}$ by it's derivative, and hence the metric is susceptible to noise even more than the derivative-based metrics.	$M_{AC} = \sum_x \sum_y [O_{x,y} \cdot O_{x,y}] - \sum_x \sum_y [O_{x,y} \cdot O_{x+1,y+1}] $

Wavelet based metrics

In addition to those presented above, there are also wavelet-transform-based focus metrics proposed in the literature [39, 61]. A major subject of those methods are the selection of the wavelet basis functions. Depending on the choice, wavelet bases possess properties such as orthogonality, symmetry, smoothness, regularity, a number of vanishing moments and a size of support [62]. After the wavelet-transform is applied to the subjected images, typically a sparsity or ratio based function is evaluated over the energies of wavelet coefficients as a focusing criteria. A variety of wavelet bases can be optimized this way to suit different applications. However, they are much more complex and time-consuming compared to the traditional focusing methods and require a priori work on the subject.

Moreover, the selection of the wavelet bases can be altered to fit better with different subjects, so they are highly object dependent. Therefore wavelet-transform-based methods are not included in the analysis.

Note that, the frequency spectrum based focus metrics can also be considered as wavelet-transform based criteria where the wavelet bases are chosen as Fourier basis functions. In those cases, the cut-off parameters are required to be manually adjusted for different images depending on their frequency spectrum [35], and hence, prior knowledge about the content of the image is required for performing reliable autofocusing.

4.3 Experimental Results on Autofocusing

The accuracy of the discussed traditional focus functions are inspected on holograms of the USAF resolution chart and a photoresist structure, which are presented in Fig. 4.2. Three phase-shifted holograms are recorded with our digital holographic microscope, and the object waves are calculated using the phase-shifting holography technique. Afterwards, the true focus distance of the holograms are estimated by the sharpness metrics. The photoresist structure is formed by coating a microscope-slide with a 2 μm height photo-resist material (AZ 5214E). Note that the edges of the photoresist structure are visible in intensity images, but it is essentially a phase object. On the other hand, the USAF chart has no phase information, and it is an example of an amplitude object.

The sharpness curves of the holograms of USAF chart and the photoresist structure are shown in figure 4.3, where some of the metrics work better than the others. For example, only the Tenenbaum gradient metric produced a high quality curve among the derivative-based focus metrics. This is due to the sensitivity of this class of functions to the speckle noise in holographic imaging. Nevertheless, since the Tenenbaum gradient metric is mathematically the most tolerant one to noise compared to the other derivative-based metrics, it performed considerably better. The histogram based focus metrics produced accurate results for both test subjects. However, the histogram range metric did not produce a reliable sharpness curve for the phase object since the range of gray-levels is small both for the in-focus and the out-of-focus images. Moreover, the histogram entropy function

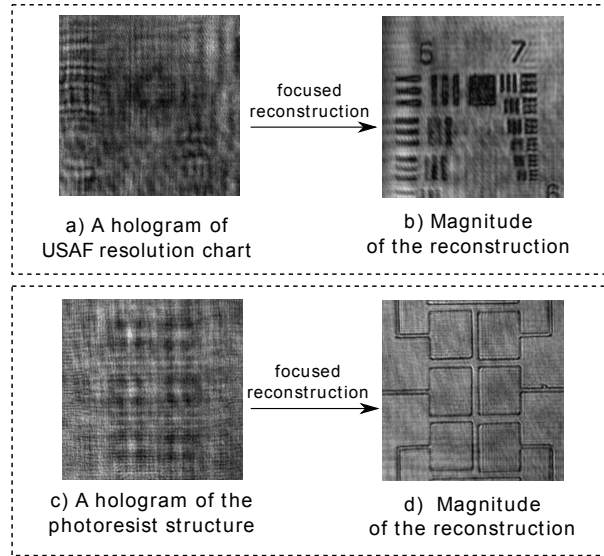


Figure 4.2: (a) and (c) are the recorded holograms of USAF resolution chart and the photoresist test subject. (b) and (d) are the reconstructions of the holograms at the exact focus distances 3 cm and 1.1 cm respectively.

surprisingly detects the true focus distance of USAF chart as a global maximum instead of a minimum. Although the USAF chart is composed of just a few histogram bins, its holographic image consists of distributed gray levels due to the speckle noise. Therefore the entropy measure is maximized when the image is focused. The sharpness curves obtained by the integral power method are also very high quality in general, except that the focus distance of the phase object is identified by a weak peak on the sharpness curve. The reason is that the object image of the phase object is not completely transparent and it has amplitude components around the structure's edges. These amplitude components combined with the speckle noise, significantly degrade the sharpness curve of the integral power metric for the photoresist structure, and therefore the peak of the sharpness curve is not very strong. The pixel count method accurately estimated the focus distance of the amplitude object but it was unable to do so for the phase object. Finally, the statistical focus measures produced accurate results for both of the object types used in the experiments. The highest quality sharpness curves are produced by the normalized-variance and the deviation-based correlation metrics. On the other hand, auto-correlation metric produced unreliable and very low-quality sharpness curves because it was greatly affected by the speckle noise.

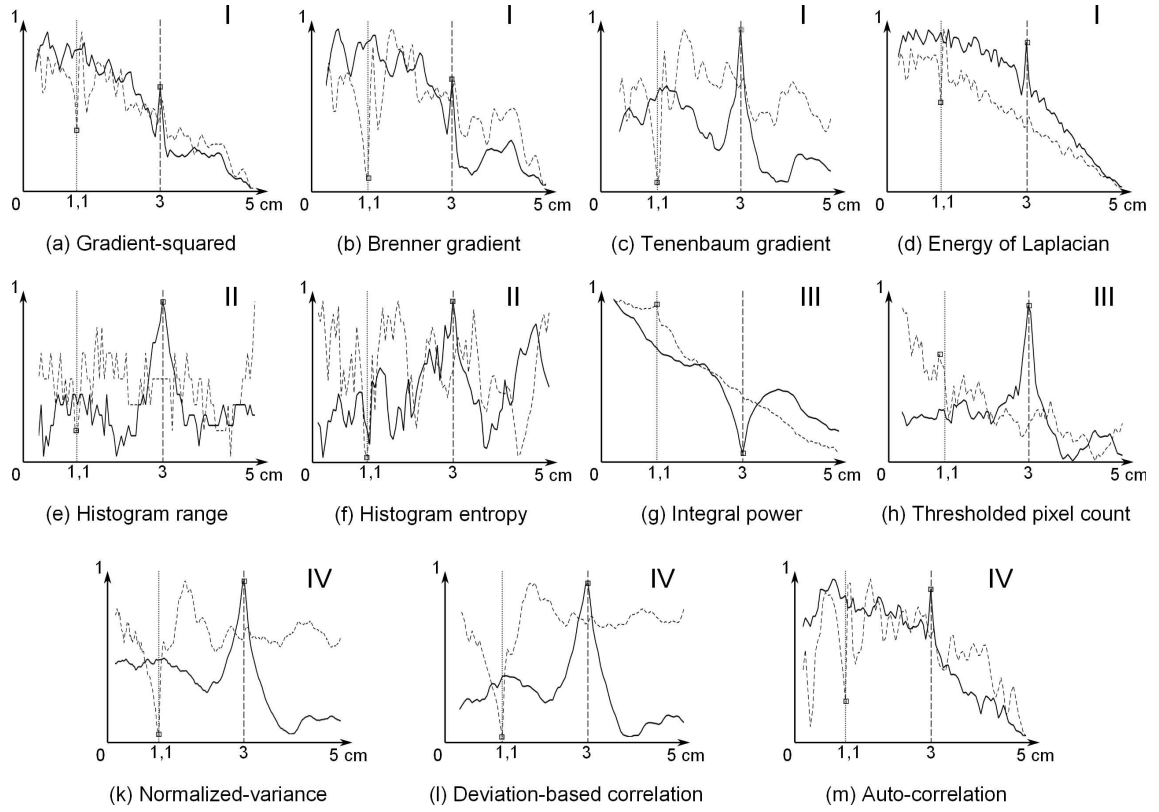


Figure 4.3: Sharpness curves obtained by traditional focus metrics using the recorded holograms of USAF resolution chart and the photoresist test subject. The true focus distances of the holograms are 3 cm and 1.1 cm respectively. The sharpness curves are calculated on a 5 cm interval with a 0.5 mm step size. The sharpness curves of the USAF chart are represented by continuous lines, and the sharpness curves of the phase object are represented by dotted lines. Tenenbaum gradient, the integral power, the normalized-variance and the deviation-based correlation metrics performed better.

In summary, all of the discussed sharpness functions produced acceptable results on the hologram of USAF resolution chart. However, the true-focus-distance of the phase object could be estimated reliably only by a minority of the sharpness functions -Tenenbaum gradient, the normalized-variance and the deviation-based correlation methods-. As it is explained above, the integral power metric is also expected to produce high quality sharpness curves for completely transparent phase objects. Therefore in overall, it is possible to say that these four metrics are the best choices as the traditional sharpness functions for estimation of the true-focus-distances in digital holography.

Chapter 5

METHODS FOR FAST AUTOFOCUSING

A real-time application typically poses a 40 ms time constraint on overall hologram processing, and it is important for an autofocusing implementation to fulfill this constraint. In the autofocusing procedure of a digital hologram, several reconstructions are performed and then a decision is made for choosing the best focused one. Regardless of the sharpness criterion, the computational bottleneck of this algorithm is the hologram reconstruction process, because it involves several two-dimensional discrete Fourier transformations (DFT). As shown in table 5.1, the sharpness estimation consumes negligible time compared to the hologram reconstruction. The complexity of reconstructing a hologram, for example, with the angular spectrum method is $\mathcal{O}(N^2 \log N)$ for a hologram of $N \times N$ pixels when the DFTs are calculated by the fast Fourier transform (FFT) method, and even a single reconstruction is very demanding in means of computational power if the holograms are large. For example a one mega-pixel sized hologram is reconstructed in approximately 800 ms, when an Intel i7-2600 CPU is used as the calculation engine [46]. Considering that several reconstructions are performed in order to find the true-focus-distance of a hologram, the autofocusing procedure becomes almost impractical. If the hologram is reconstructed at M candidate distances, then the complexity of the autofocusing procedure increases to $\mathcal{O}(MN^2 \log N)$. To decrease the computational load of autofocusing, one can resort to methods such as using a more intelligent interval search

algorithm or utilizing powerful parallel processing hardwares as the calculation engine. These methods improve the performance, since the former reduces the number of reconstructions required for performing autofocusing and the latter increases the calculation speed of reconstructions. In addition to these methods, we also proposed a scaling technique that is applied on holograms before reconstruction [46]. In this chapter, these three methods for increasing the speed performance of autofocusing are explained, and the scaling technique is investigated.

Table 5.1: The time consumed by a hologram reconstruction using angular spectrum method and the time consumed by a sharpness estimation using normalized-variance and gradient-squared metrics are shown. Hologram reconstruction is completed much more slower than the sharpness estimation, and thus it is the bottleneck in an autofocusing algorithm. Note that the timings are obtained with an Intel i7-2600 CPU working on a single core.

Hologram size	Angular Spect.	Norm.-Var.	Grad.-Sq.
1024×1024	813 ms	21.1 ms	42 ms

5.1 Intelligent Search Algorithms

Instead of performing a linear search on an interval, one can use a more intelligent search algorithm to reduce the number of reconstructions required for performing autofocusing. In an autofocusing procedure with linear search algorithm, as implemented in this thesis, an interval is divided by a fixed step-size, and each step is regarded as a candidate true-focus-distance where a hologram reconstruction operation and a sharpness estimation is performed at each step. This procedure is highly inefficient. Other than the linear search, one can resort to search algorithms such as the iterative search, the variable step-size search, the binary search or the Fibonacci search [63, 64]. Using these search algorithms, it is possible to find the peak of the sharpness curves without testing each propagation distance. However, the termination condition of these intelligent search algorithms define the accuracy of the estimations. It is reported that Fibonacci search outperforms other

search algorithms in autofocusing of digital holograms [65].

5.2 Utilizing the Graphics Processing Unit (GPU)

Another way to increase the speed of a hologram reconstruction is using a more powerful calculation engine. Unless coded specifically, computers use the central processing unit (CPU) while executing programs. State-of-the-art CPUs can perform up to a hundred billion floating operations per second (GFLOPS) sequentially, and this calculation speed is sufficient most of the time. However high complexity operations such as hologram reconstruction takes hundreds of milliseconds when the hologram size is large. To increase the performance, the reconstruction operation can be distributed to multiple processing units. The graphics card of a computer offers a good alternative for executing parallel computations. As shown in table 5.2, one can achieve up to a approximately hundred times speed-up in reconstructions using GPUs as the calculation engine instead of CPUs [42].

Table 5.2: Reconstruction times in milliseconds for the hologram reconstruction methods. 1024×1024 holograms are reconstructed by an Intel XEON W3670 CPU and an Nvidia GTX 660 Ti graphics card.

Reconstruction method	CPU (ms)	GPU (ms)	Speed-up
Fresnel approximation	937	9.2	$101.8 \times$
Convolution	734	9	$81.5 \times$
Angular Spectrum	941	9.1	$103.4 \times$

Although it is possible to increase the performance of autofocusing by utilizing GPUs, this comes with a capital cost for the hardware and a development cost of its integration to the holographic setup. Instead, one can use the scaling technique as it is explained below to lower the computation time of autofocusing without adding more resources to the system.

5.3 Autofocusing using The Scaling Technique

As the complexity of the autofocusing procedure is $\mathcal{O}(MN^2 \log N)$ for an image with $N \times N$ pixels, a reduction in the hologram size would imply a significant increase in the performance of the algorithm. It is shown that scaled-down versions of a hologram may

Table 5.3: The figure shows the execution time of a single hologram reconstruction using the convolution method and a single execution of the normalized-variance metric. The timings are obtained on an Intel i7-2600 CPU.

Hologram size	Convolution	Normalized-Variance	Total
1024×1024	684 ms	21.1 ms	705.1 ms
512×512	174 ms	5.2 ms	181.2 ms
256×256	45.4 ms	2.9 ms	48.3 ms
128×128	12.1 ms	0.8 ms	12.9 ms
64×64	3.45 ms	0.22 ms	3.67 ms
32×32	1.2ms	0.07 ms	1.27 ms

be used to estimate the true-focus-distance of the original hologram up to a certain scale ratio [66]. Scaling operation with ratio k is defined as the division of the hologram in to $k \times k$ pixels squares, and then using the average intensity and phase values of each square as the pixel values of the new scaled hologram as demonstrated in figure 5.1. Since this new hologram is composed of $\frac{N}{k} \times \frac{N}{k}$ pixels, its numerical reconstruction will be completed approximately $k^2 \log k$ times faster than the original. Figure 5.3 shows the speed-up achieved by using scaled holograms while performing a reconstruction and assessing its sharpness. The pixel dimensions of the scaled hologram is also changed and the new pixel dimensions are now k times larger than the original, that is:

$$\Delta x_s = k\Delta x, \quad \Delta y_s = k\Delta y, \quad (5.1)$$

where Δx_s and Δy_s are the pixel dimensions of the scaled hologram, and Δx and Δy are the pixel dimensions of the original hologram.

To perform focusing using the scaling technique, a recorded hologram is scaled-down with ratio k . Then, the sharpness curve of this scaled hologram is calculated where the peak of the curve indicates the true-focus-distance. Afterwards, the original hologram is reconstructed at the true-focus-distance and a focused image of the sample is achieved. An exemplary focusing operation using scaling is demonstrated on a hologram of USAF chart in figure 5.2.

While it is possible to find the true-focus-distance of a hologram using the sharpness curve of its scaled-down version, the maximum scaling limit without diverging from the original sharpness curve is not obvious and there is a limit. The sharpness curve of a hologram starts to broaden and its peak disappears as the scale-ratio k is increased, because the scaling operation has a low-pass characteristic. The maximum limit for the scale-ratio is dependent on the frequency content of the hologram, and it can be increased to the point which the high frequency content in the image is significantly lost and the resultant sharpness curves no longer approximates the original. Holograms of objects with balanced spatial frequency spectrum can be scaled more without diverging from the sharpness curve of the original hologram. The exact scaling limit depends on the hologram, but it is shown that eight times scaling on mega-pixel size holograms provide robust and reliable sharpness curves [45]. On the other hand, for example, the mega-pixel size hologram of USAF chart can be scaled 32 times without losing reliability as shown in fig-

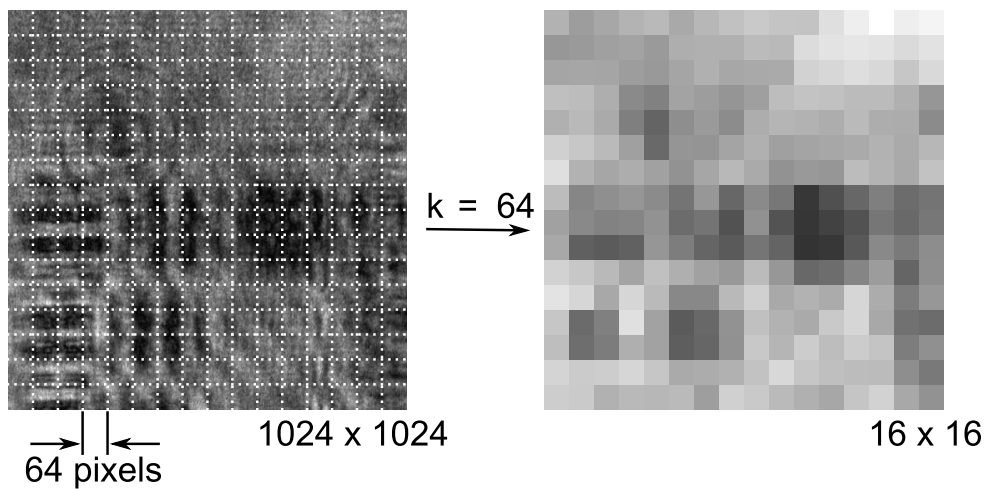


Figure 5.1: Scaling operation is demonstrated. A hologram is divided into k by k squares, and the average intensity of each square is set as the corresponding pixel value of the new scaled hologram.

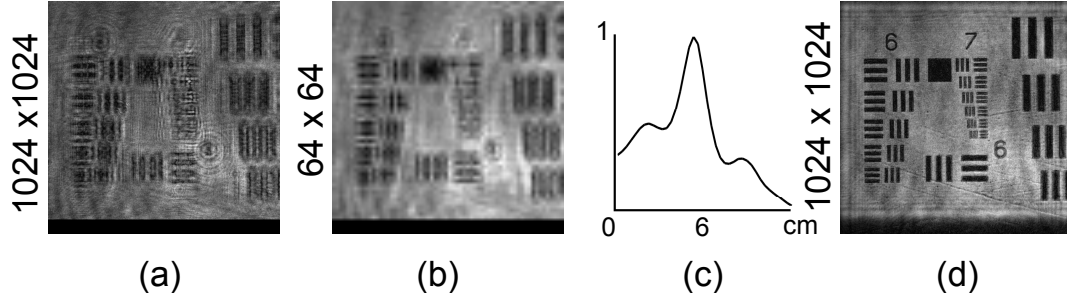


Figure 5.2: An exemplary autofocus using the scaling technique. A recorded hologram of the USAF chart (a) is scaled with ratio $k = 16$ and the scaled-hologram (b) is obtained. Then the sharpness curve of this new scaled hologram (c) is computed using the normalized-variance and the gradient-squared metrics, represented by continuous and dotted lines respectively. The true-focus-distance of the scaled hologram is identified as 3 cm by both of the sharpness metrics as a global maxima. Finally, the original sized hologram is reconstructed at true-focus-distance and a focused image of the object is achieved (d).

ure 5.3, where approximately 1000 times speed-up is achieved. Scaling is an easy way to decrease the computational load of the autofocusing procedure, and it will gain more importance in the future considering the increasing rate of the pixel density of CCD cameras in todays development pace.

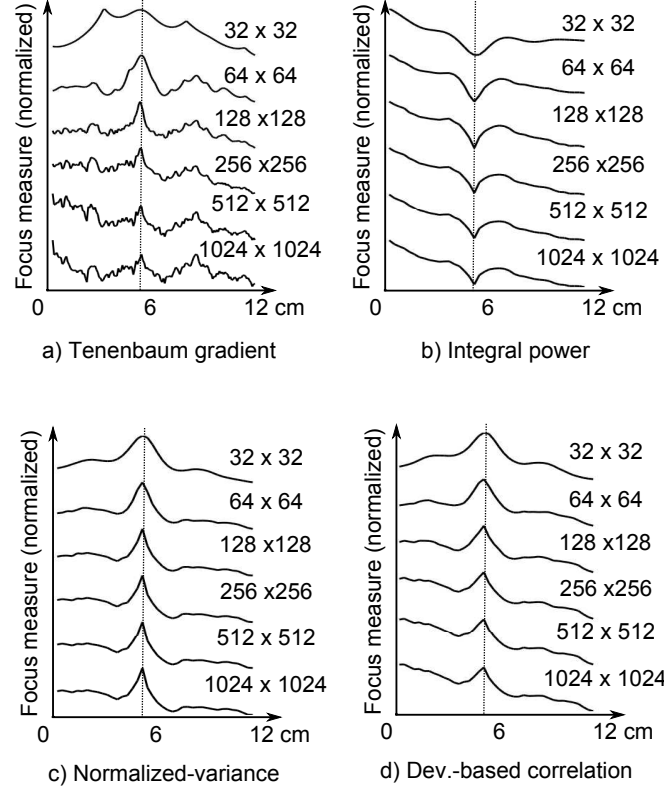


Figure 5.3: The effect of scaling on sharpness curves is shown on a recorded hologram of USAF resolution chart. Hologram is originally composed of 1024×1024 pixels. Two, four, eight, 16 and 32 times scaling is applied and the sharpness curves are calculated by the four chosen metrics. As k , increases curves mostly become wider and smoother due to the low-pass nature of the scaling.

5.3.1 Autofocusing Experiments using the Scaling Technique

The scaling technique increases the speed of autofocusing greatly, however this operation also degrades the sharpness curve. In this section, breakdown limits of autofocusing using scaled holograms is tested in an experimental situation. Human cheek epithelial cells are collected with a cotton swab and placed onto a microscope slide. The microscope slide is inserted into the digital holographic microscope, and three phase-shifted holograms are captured at an unknown focus distance. After performing the phase-shifting technique, the focus distance of the cells are estimated by applying the four different sharpness metrics (the Tenenbaum gradient, the integral power, the normalized-variance and the deviation-based correlation) on the magnitude of the reconstructed holograms. The reason for choosing those four sharpness metrics is that they have performed very well for both object types, and they could also preserve accuracy even with very high

ratios of scaling. Only the Tenenbaum gradient metric among the derivative-based sharpness metrics is included in the analysis, because these metrics use the same principle, but the Tenenbaum gradient metric has the most tolerance to noise. On the other hand, histogram-based metrics are discluded from the analysis because they have a poor performance under high amounts of scaling. The reason for this is that the averaging effect of scaling groups histogram bins towards the mean gray-level bin, and therefore these metrics lose accuracy greatly as the images are scaled. Thresholded pixel count metric is also excluded because it does not work for phase objects. Finally, the auto-correlation metric is not included in the analysis, because it is highly affected by noise (even more than the derivative based metrics) and it produces low quality curves even with small amounts of scaling.

Although the specimen was transparent, the four chosen focus metrics were able to estimate the true-focus-distance correctly. Fig. 5.4-a and Fig. 5.4-b shows the reconstructed magnitude and phase images of the specimen at the true-focus-distance. Note that the epithelial cells are not visible in the focused magnitude image, but they can be seen very well in detail in the phase image. It is also worth to note that, since the sample is completely invisible when on focus, integral power metric performed very well and produced reliable sharpness curves.

In order to measure the degradation in the performance of metrics under scaling, two measures are defined for comparing sharpness curves. Firstly, the quality measure is defined as the inverse of full width at half-maximum (FWHM) of the curve around the global extrema, and secondly, the deviation measure is defined as the deviation of the global extrema from the true focus point.

Epithelial cell hologram is scaled by increasing scale-ratios, and its focus distance is estimated by each of the four metrics. Then, the quality and the deviation of the sharpness curves are computed. The results are shown in Fig. 5.4-c and Fig. 5.4-d as a function of the scale-ratio [41]. Surprisingly when the image is scaled by two-folds for the first time, the sharpness curve quality of the Tenenbaum gradient metric increases. This is due to the low-pass characteristic of the scaling method that filters out high frequency noises which are the speckle noise and the differentiation noise in this case. Nevertheless, the

Tenenbaum gradient metric's sharpness curve continues to degrade just in the others when the scale-ratio is increased more. If the epithelial cell hologram was a noise-free one, the measured sharpness curve quality for the original image size would be higher, and it would continuously degrade as the image was scaled. Another important observation is that 8 times scaling can be applied for autofocusing 1024×1024 pixels holograms of human cheek epithelial cells with less than 1% divergence from the true-focus-distance. With 8 times scaling, approximately 64 times speed-up is achieved in the autofocusing process without losing reliability. This is a satisfactory improvement considering that the overall speed-up would be over five thousand times if GPUs were to be utilized in addition [45].

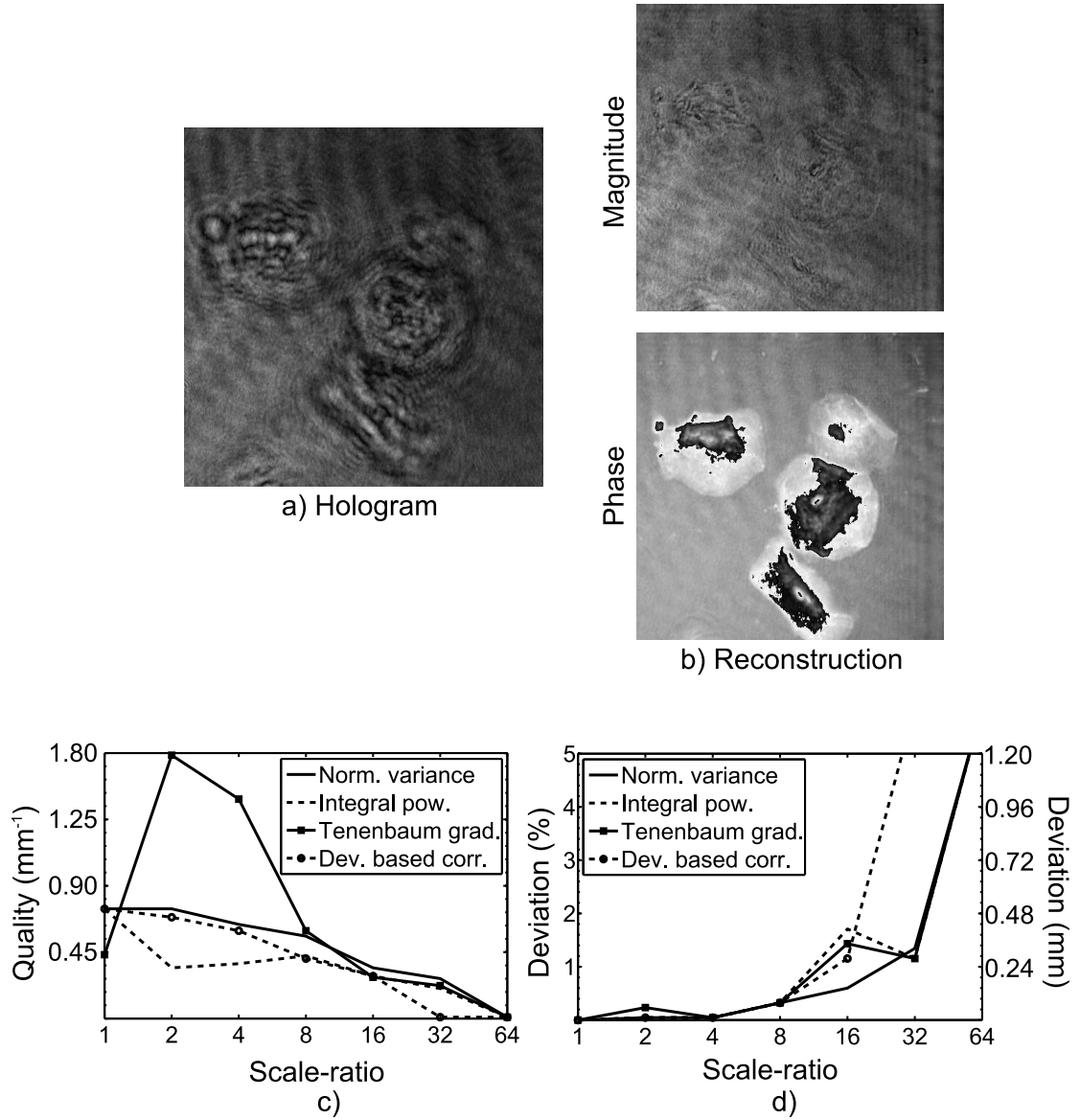


Figure 5.4: (a) Hologram of human cheek epithelial cells, and (b) its reconstruction at the true focus distance. The magnitude of the reconstructed holograms are used for finding the true focus distance (2.4 cm in this case), where the epithelial cells are normally invisible in the image when it is focused (b - top). After finding the focus distance, unstained epithelial cells are visible in detail in the phase image of the reconstruction (b - bottom). To find the focus distance, the hologram is scaled two, four, eight, 16 and 32 times, and the sharpness curve of each scaled version is computed using the normalized-variance, the integral power, Tenenbaum gradient and the deviation-based correlation metrics. The change in the resolution and the peak-point deviation of the sharpness curves obtained by these four methods are shown in (c) and (d) respectively for increasing scale-ratios. As the resolution of the curves decrease, it is possible to scale up to 8 times safely without diverging from the actual focus distance

5.3.2 Statistical Investigation of the Scaling Technique on Computer Generated Holograms

The scaling method increases the speed of hologram reconstructions greatly since the number of pixels decreases. However, it is not obvious how the sharpness curves are affected by the scaling operation. It is also not clear how much the scale-ratio can be increased while preserving acceptable quality and accuracy on the sharpness curves, because the exact scaling limit is dependent on the hologram content. To be able to perform autofocusing using scaling in an unsupervised manner, it is required to know how much scaling can be applied regardless of the hologram content. For that reason, it is essential to measure the degradation in the quality and the accuracy of the sharpness curves statistically as a function of the scale-ratio.

Fifty different computer generated holograms of 1024×1024 pixels gray-scale images of human portraits, test-structures, landscapes and micro-particles are used for testing the affects of scaling on autofocusing of digital holograms. These images are listed in figure 5.5. The holograms are generated 25 cm away from the camera plane, and then they are scaled using the scaling technique with scale-ratios two, four, eight, 16, 32 and 64. Afterwards, the scaled holograms are reconstructed in 100 steps in an interval from 22 cm to 28 cm. Then for each scale-ratio, the sharpness curves are calculated by the four previously chosen metrics, which are the normalized-variance, the integral power, the deviation-based correlation and Tenenbaum gradient metrics [41].

For each metric and the scale-ratio, the calculated sharpness curves are overlaid on top of each other in figure 5.6. As expected, the sharpness curves get wider and smoother as the scale-ratio is increased due to the low-pass effect of this operation. However, the results show that sharpness curves of scaled holograms may be used for estimating the true-focus-distance because the peak of the sharpness curves remain consistent as the scale-ratio is increased. It appears that for 1024×1024 pixels images, the peak of the sharpness curves start to shift starting from the scale-ratio 32 (image size 32×32). Another important observation is that the Tenenbaum gradient metric works better and better as the scale-ratio is increased. Compared to the other metrics, this metric produced visually the most clean sharpness curves with accurate peak points when scale-ratio is set

to 16. In overall, all four of the metrics work well with the scaling method. However it is worth to note that the deviation based correlation method appears to break down sooner than the other three methods. The sharpness curves of this metric are severely degraded for scale-ratios greater than 8, where the other metrics appear to sustain scale-ratio 16 very well.

To quantitatively evaluate the degradation of the sharpness curves, the quality and the deviation measures are defined as the inverse of full width at half-maximum (FWHM) of the curve around the peak, and the deviation of the peak from the true-focus-point respectively. In Fig. 5.7-a and 5.7-b, the quality and the deviation of the sharpness curves of each metric are averaged, and the average values are represented as a function of the scale-ratio. It is observed that the average quality of the sharpness curves decrease, and the average deviation on the peak point of the sharpness curves increase as expected when the scale-ratio is increased. This is, again, due to the low-pass characteristic of the scaling operation in which the cut-off frequency is proportional to the scale-ratio. Even so, it is seen from figure 5.7-b that the deviation from the true-focus-distance is still less than 2% when 16 times scaling is applied. For the scale-ratio eight, the deviation of all of the four metrics are on the order of one millimeter. As it was observed earlier from the overlaid sharpness curves, the deviation based correlation metric diverges sooner than the other metrics for scale-ratios greater than 8. Nevertheless, scaling 1024×1024 pixels holograms more than 16 times is not practical because the deviation rises much faster. It is also worth noting that the integral power metric appears to have the most tolerance to scaling among the four chosen metrics where the deviation is remained below 2% until 64 times scaling. We suspect that since the integral power metric only sums up the pixel values, this metric is already similar to low-pass filtering, and therefore it is inherently the most resistant one to the low-pass effect of scaling compared to the other sharpness metrics.



Figure 5.5: The list of fifty different 1024×1024 pixels gray-scale images that are used for creating computer generated holograms.

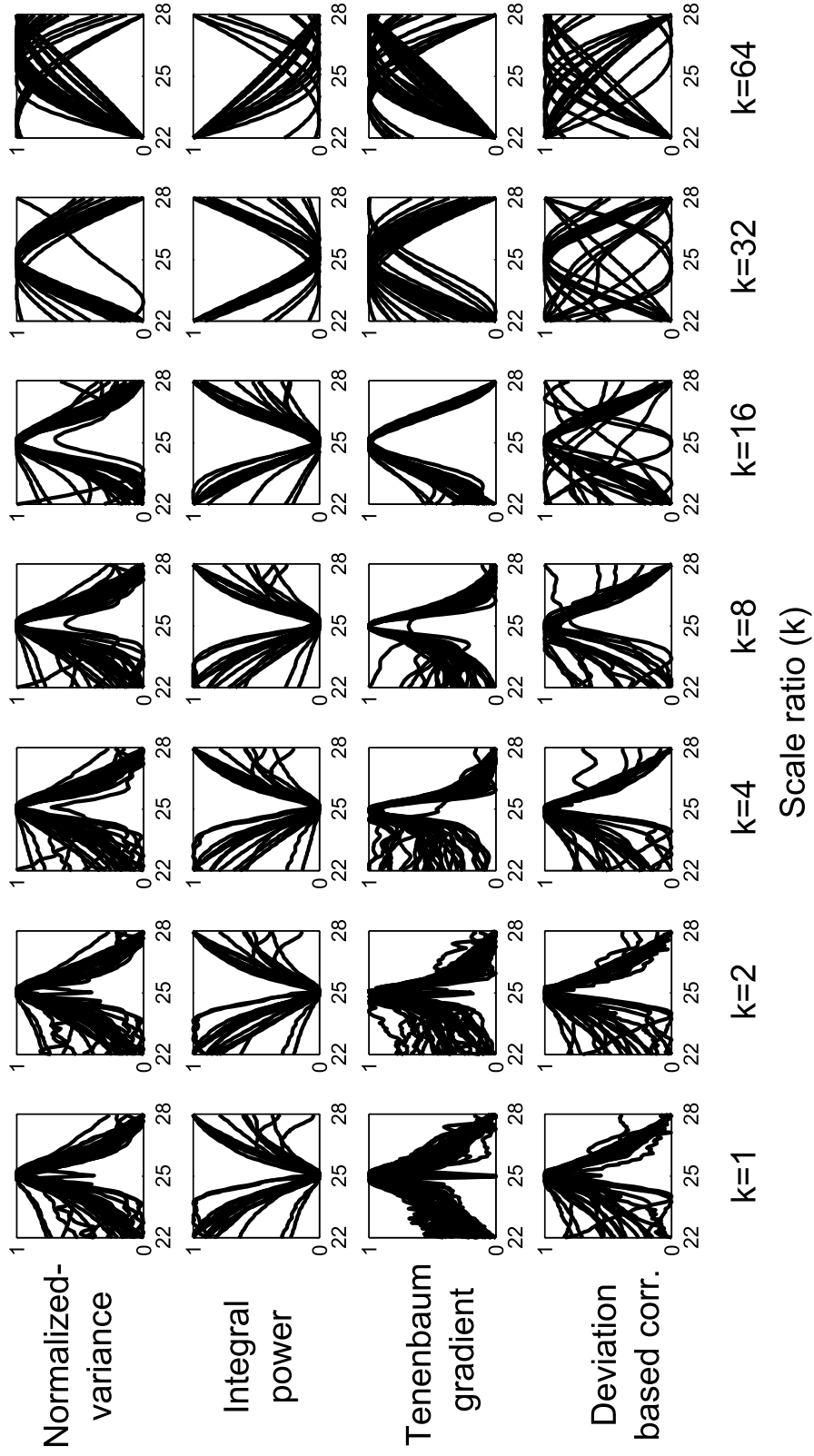


Figure 5.6: Fifty different computer generated holograms of 1024×1024 pixels are subjected to sharpness estimation using the normalized-variance, the integral power, Tenenbaum gradient and the deviation-based correlation metrics. For each scale-ratio, computer generated holograms are scaled, and then the sharpness curves are calculated by each method in 100 steps between 22 cm and 28 cm. For each metric, the curves obtained from fifty images are overlaid on each other as a function of the scale-ratio.

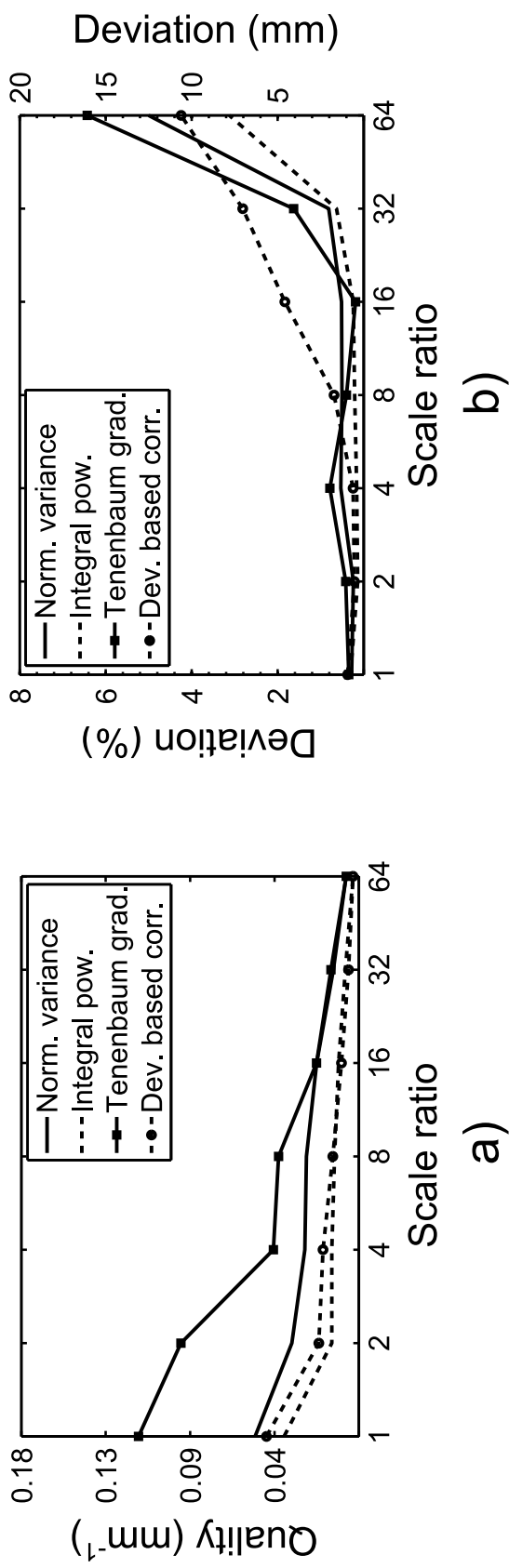


Figure 5.7: Fifty different computer generated holograms of 1024×1024 pixels are subjected to sharpness estimation using the normalized-variance, the integral power, Tenenbaum gradient and the deviation-based correlation metrics. For each scale-ratio, computer generated holograms are scaled, and then the sharpness curves are calculated by each method in 100 steps. (a) Average sharpness curve quality (FWHM^{-1} of the global extrema) for each metric as a function of scale-ratio. (b) Average deviation of the global extrema from the true focus distance for each metric as a function of scale-ratio. Note that the normalized-variance and the deviation-based correlation metrics produced some flipped and unreliable sharpness curves when $k \geq 16$. For example, for $k = 32$ a single curve of the normalized-variance metric is flipped. Those flipped curves are discarded while calculating the average performance of the metrics.

Chapter 6

CONCLUSION

In this thesis, autofocusing of digital holograms is investigated using common sharpness functions developed for standard photography and microscopy. Holograms are recorded with a digital holographic microscope, and a traditional autofocusing algorithm is executed for estimating the optimal reconstruction distance. Eleven of the most common sharpness functions are tested on amplitude and phase objects, where the normalized-variance, the integral power, Tenenbaum gradient and the deviation-based correlation methods are observed to be the most reliable and robust focus metrics. Moreover, it is shown that autofocusing of mega-pixels size holograms is completed in almost a second, and optimization is required for operation in real-time. A scaling technique is proposed for increasing the performance of autofocusing in means of execution time, where the speed of a reconstruction is improved on the order of square of the scale-ratio. The scaling technique is applied during the autofocusing of a hologram of a biological specimen, and successful results are obtained, that is, 64 times speed-up is achieved with less than 1% deviation from the true focus distance. The scaling technique is also applied to a hologram of USAF resolution chart where approximately 1000 times speed-up is achieved in autofocusing. Since the scaling technique detracts the quality of the sharpness curves, the technique is statistically investigated on fifty computer generated holograms of various images. It is shown that the deviation from the true-focus distance of the original holograms is less than 2% for scale-ratios lesser than or equal to 16. It is worth to note that the speed improvement is independent of the processing power of the system, and it

could be further improved by approximately 100 times if GPUs were to be utilized.

References

- [1] D. Gabor, "A new microscopic principle," *Nature*, vol. 161, pp. 777–778, 1948.
- [2] R. G. Gould, "The laser, light amplification by stimulated emission of radiation," in *The Ann Arbor Conference on Optical Pumping* (P. A. Franken and R. H. Sands, eds.), p. 128, 1959.
- [3] E. N. Leith and J. Upatnieks, "Reconstructed wavefronts and communication theory," *Journal of the Optical Society of America*, vol. 52, no. 10, pp. 1123–1130, 1962.
- [4] R. L. Powell and K. A. Stetson, "Interferometric vibration analysis by wavefront reconstructions," *Journal of the Optical Society of America*, vol. 55, pp. 1593–1598, 1965.
- [5] K. A. Stetson and R. L. Powell, "Interferometric hologram evaluation and real-time vibration analysis of diffuse objects," *Journal of the Optical Society of America*, vol. 55, pp. 1694–1695, 1965.
- [6] A. Macovski, "Time-lapse interferometry and contouring using television systems," *Applied Optics*, vol. 10, no. 12, pp. 2722–2727, 1971.
- [7] B. R. Brown and A. W. Lohmann, "Computer-generated binary holograms," *IBM Journal of Research and Development*, vol. 13, pp. 160–168, 1969.
- [8] W. H. Lee, "Computer-generated holograms: techniques and applications," *Progress in Optics*, vol. 16, pp. 120–232, 1978.
- [9] O. Bryngdahl and F. Wyrowski, "Digital holography - computer generated holograms," *Progress in Optics*, vol. 28, pp. 1–86, 1990.

- [10] T. S. Huang, "Digital holography," in *Proceedings of the IEEE*, vol. 59, pp. 1335–1346, 1971.
- [11] M. A. Kronrod, L. P. Yaroslavski, and N. S. Merzlyakov, "Computer synthesis of transparency holograms," *Soviet Physics-Technical Physics*, vol. 17, pp. 329–332, 1972.
- [12] M. A. Kronrod, N. S. Merzlyakov, and L. P. Yaroslavski, "Reconstruction of holograms with a computer," *Soviet Physics-Technical Physics*, vol. 17, pp. 333–334, 1972.
- [13] J. W. Goodman and R. W. Lawrence, "Digital image formation from electronically detected holograms," *Applied Physics Letters*, vol. 11, pp. 777–778, 1967.
- [14] U. Schnars and W. Jüptner, "Principles of direct holography for interferometry," in *FRINGE 93: Proceedings of 2nd International Workshop on Automatic Processing of Fringe Patterns* (W. Jüptner and W. Osten, eds.), pp. 115–120, 1993.
- [15] U. Schnars and W. P. O. Jüptner, "Digital recording and numerical reconstruction of holograms," *Measurement Science and Technology*, vol. 13, pp. R85–R101, 2002.
- [16] E. Kolenovic, S. Lai, and W. Osten, "Endoscopic shape and deformation measurement by means of digital holography," in *Proceeding of 4th International Workshop on Automatic Processing of Fringe Patterns* (W. Jüptner and W. Osten, eds.), pp. 686–691, 2001.
- [17] W. Osten, S. Seebacher, T. Baumbach, and W. Jüptner, "Absolute shape control of microcomponents using digital holography and multiwavelength contouring," in *Proceedings of SPIE*, vol. 4275, pp. 71–84, 2001.
- [18] W. Haddad, D. Cullen, J. C. Solem, J. M. Longworth, A. McPherson, K. Boyer, and C. K. Rhodes, "Fourier-transform holographic microscope," *Applied Optics*, vol. 31, pp. 4973–4978, 1992.

- [19] L. Xu, X. Peng, J. Miao, and A. K. Asundi, “Studies of digital microscopic holography with applications to microstructure testing,” *Applied Optics*, vol. 40, pp. 5046–5051, 2001.
- [20] V. Kebbel, H. J. Hartmann, and W. Jüptner, “Application of digital holographic microscopy for inspection of micro-optical components,” in *Proceedings of SPIE*, vol. 4398, pp. 189–198, 2001.
- [21] Y. Emery, E. Cuhe, T. Colomb, C. Depeursinge, B. Rappaz, P. Marquet, and P. Magistretti, “Dhm (digital holographic microscope) for imaging cells,” *Journal of Physics: Conference Series*, vol. 61, pp. 1317–1321, 2007.
- [22] M. Adams, T. Kreis, and W. Jüptner, “Particle size and position measurement with digital holography,” in *Proceedings of SPIE*, vol. 3098, pp. 234–240, 1997.
- [23] R. B. Owen and A. A. Zozulya, “In-line digital holographic sensor for monitoring and characterizing marine particulates,” *Optical Engineering*, vol. 39, pp. 2187–2197, 2000.
- [24] C. B. Lefebvre, S. Coëtmellec, D. Lebrun, and C. Özkul, “Application of wavelet transform to hologram analysis: three-dimensional location of particles,” *Optics and Lasers in Engineering*, vol. 33, pp. 409–421, 2000.
- [25] R. Dandliker and K. Weiss, “Reconstruction of the three-dimensional refractive index from scattered waves,” *Optics Communications*, vol. 1, pp. 323–328, 1970.
- [26] V. Kebbel, M. Adams, H. J. Hartmann, and W. Jüptner, “Digital holography as a versatile optical diagnostic method for microgravity experiments,” *Measurement Science and Technology*, vol. 10, pp. 893–899, 1999.
- [27] L. Hesselink, S. S. Orlov, and M. C. Bashaw, “Holographic data storage systems,” in *Proceedings of the IEEE*, vol. 92, pp. 1231–1280, 2004.
- [28] J. Ashley, M. P. Bernal, G. W. Burr, H. Coufal, H. Guenther, J. A. Hoffnagle, C. M. Jefferson, B. Marcus, R. M. Macfarlane, R. M. Shelby, and G. T. Sincerbox, “Holo-

- graphic data storage,” *IBM Journal of Research and Development*, vol. 44, no. 3, pp. 341–368, 2000.
- [29] J. Zhai, G. Jin, Y. Ruan, Y. Yan, and M. Wu, “Study on data storage format of a holographic optical disk,” *Optics Communications*, vol. 143, pp. 185–188, 1997.
- [30] Y. Aoki, “Watermarking technique using computer-generated holograms,” *Electronics and Communications in Japan, Part 3*, vol. 84, no. 1, pp. 21–31, 2001.
- [31] J. W. Goodman, *Introduction to Fourier Optics*. New York: McGraw–Hill Companies, Inc., second ed., 1996.
- [32] I. Sexton and P. Surman, “Stereoscopic and autostereoscopic display systems,” *IEEE Signal Processing Magazine*, vol. 16, no. 3, pp. 85–99, 1999.
- [33] M. Kempkes, T. Vetter, and M. Mazzotti, “Measurement of 3d particle size distribution by stereoscopic imaging,” *Chemical Engineering Science*, vol. 65, pp. 1362–1373, 2010.
- [34] H. Arimoto and B. Javidi, “Integral three-dimensional imaging with digital reconstruction,” *Optics Letters*, vol. 26, pp. 157–159, 2001.
- [35] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston, “Comparison of autofocus methods for automated microscopy,” *Cytometry*, vol. 12, pp. 195–206, 1991.
- [36] F. C. Groen, I. T. Young, and G. Ligthart, “A comparison of different focus functions for use in autofocus algorithms,” *Cytometry*, vol. 6, pp. 81–91, 1985.
- [37] Y. Sun, S. Duthaler, and B. J. Nelson, “Autofocusing in computer microscopy: selecting the optimal focus algorithm,” *Microscopy Research and Technique*, vol. 65, pp. 139–149, 2004.
- [38] J. Gillespie and R. King, “The use of self-entropy as a focus measure in digital holography,” *Pattern Recognition Letters*, vol. 9, no. 1, pp. 19 – 25, 1989.

- [39] M. Liebling and M. Unser, “Autofocus for digital fresnel holograms by use of a fresnelet-sparsity criterion,” *Journal of the Optical Society of America*, vol. 21, pp. 2424–3062, 2004.
- [40] F. Dubois, C. Schockaert, N. Callens, and C. Yourassowsky, “Focus plane detection criteria in digital holography microscopy by amplitude analysis,” *Optics Express*, vol. 14, pp. 5895–5908, 2006.
- [41] H. A. İlhan, M. Doğar, and M. Özcan, “Digital holographic microscopy and focusing methods based on image sharpness,” *Journal of Microscopy*, vol. 255, pp. 138–149, 2014.
- [42] M. Doğar, H. A. İlhan, and M. Özcan, “Realtime reconstruction of digital holograms with gpu,” in *Proceedings of SPIE 8644, Practical Holography XXVII: Materials and Applications*, vol. 86440B-8, 2013.
- [43] T. Shimobaba, Y. Sato, J. Miura, M. Takenouchi, and T. Ito, “Real-time digital holographic microscopy using the graphic processing unit,” *Optics Express*, vol. 16, pp. 11776–11781, 2008.
- [44] L. Ahrenberg, A. J. Page, B. M. Hennelly, J. B. McDonald, and T. J. Naughton, “Using commodity graphics hardware for real-time digital hologram view-reconstruction,” *Journal of Display Technology*, vol. 5, pp. 111–119, 2009.
- [45] M. Doğar, H. A. İlhan, and M. Özcan, “Real-time, auto-focusing digital holographic microscope using graphics processors,” *Review of Scientific Instruments*, vol. 84, pp. 083704–083704–7, 2013.
- [46] H. A. İlhan, M. Doğar, and M. Özcan, “Fast autofocusing in digital holography using scaled holograms,” *Optics Communications*, vol. 287, pp. 81–84, 2013.
- [47] U. Schnars and W. P. O. Jüptner, “Digital recording of holograms by a ccd target and numerical reconstruction,” *Applied Optics*, vol. 33, no. 2, pp. 179–181, 1994.
- [48] M. Bayraktar, *Digital Holography and Hybrid Opto-Acoustic Imaging System for Vibration Analysis*. PhD thesis, Sabanci University, 2010.

- [49] Y. Takaki, H. Kawai, and H. Ohzu, “Hybrid holographic microscopy free of conjugate and zero-order images,” *Applied Optics*, vol. 38, pp. 4990–4996, 1999.
- [50] M. K. Kim., *Digital Holographic Microscopy: Principles, Techniques, and Applications*. Springer, 2011.
- [51] M. Özcan and M. Bayraktar, “Digital holography image reconstruction methods,” in *Proceedings of SPIE 7233, Practical Holography XXIII: Materials and Applications*, vol. 72330B-1, 2009.
- [52] D. B. Kirk and W. M. W. Hwu, *Programming Massively Parallel Processors: A Hands-On Approach*. Morgan Kaufmann, second ed., 2012.
- [53] R. A. Muller and A. Buffington, “Real-time correction of atmospherically degraded telescope images through image sharpening,” *Journal of the Optical Society of America*, vol. 64, pp. 1200–1210, 1974.
- [54] J.-M. Geusebroek, F. Cornelissen, A. W. M. Smeulders, and H. Geerts, “Robust autofocusing in microscopy,” *Cytometry*, vol. 39, pp. 1–9, 2000.
- [55] P. Langehanenberg, B. Kemper, D. Dirksen, and G. von Bally, “Autofocusing in digital holographic phase contrast microscopy on pure phase objects for live cell imaging,” *Applied Optics*, vol. 47, no. 19, pp. 176–182, 2008.
- [56] P. Memmolo, C. Distanto, M. Paturzo, A. Finizio, P. Ferraro, and B. Javidi, “Automatic focusing in digital holography and its application to stretched holograms,” *Opt Lett.*, vol. 36, pp. 1945 – 1947, 2011.
- [57] W. Li, N. C. Loomis, Q. Hu, and C. S. Davis, “Focus detection from in-line holograms based on spectral 11 norms,” *Journal of the Optical Society of America*, vol. 24, pp. 3054–3062, 2007.
- [58] J. F. Brenner, B. S. Dew, J. B. Horton, T. King, P. W. Neurath, and W. D. Selles, “An automated microscope for cytologic research a preliminary evaluation,” *Journal of Histochemistry and Cytochemistry*, vol. 24, no. 1, pp. 100–111, 1976.

- [59] T. T. E. Yeo, S. H. Ong, Jayasooriah, and R. Sinniah, “Autofocusing for tissue microscopy,” *Image and Vision Computing*, vol. 11, pp. 629–639, 1993.
- [60] D. Vollath, “Automatic focusing by correlative methods,” *Journal of Microscopy*, vol. 147, pp. 279–288, 1987.
- [61] G. Yang and B. J. Nelson, “Wavelet-based autofocusing and unsupervised segmentation of microscopic images,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, pp. 2143–2148, 2003.
- [62] S. Mallat, *A Wavelet Tour of Signal Processing*. Massachusetts: Elsevier Inc., third ed., 2009.
- [63] C. F. Batten, “Autofocusing and astigmatism correction in the scanning electron microscope,” Master’s thesis, University of Cambridge, 2000.
- [64] D. E. Ferguson, “Fibonacci searching,” *Communications of the ACM*, vol. 3, p. 648, 1960.
- [65] C. M. Elhinney, *Digital Hologram Image Processing*. PhD thesis, National University of Ireland, 2009.
- [66] H. A. İlhan, M. Doğar, and M. Özcan, “Autofocusing in digital holography,” in *Proceedings of SPIE 8644, Practical Holography XXVII: Materials and Applications*, vol. 86440C-11, 2013.